LAYERED DEFORMOTION WITH RADIANCE: A MODEL FOR APPEARANCE, SEGMENTATION, REGISTRATION, AND TRACKING

A Thesis
Presented to
The Academic Faculty

by

Jeremy D. Jackson

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2007

LAYERED DEFORMOTION WITH RADIANCE: A MODEL FOR APPEARANCE, SEGMENTATION, REGISTRATION, AND TRACKING

Approved by:

Dr. Anthony J. Yezzi, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Aaron Lanterman
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Allen Tannenbaum
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Greg Turk
College of Computing
Georgia Institute of Technology

Dr. Patricio Vela
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: 3 July 2007

Dedicated to my parents,

Dr. Thomas and Louise Jackson,

my brothers, Jonathan and Joshua,

and my sister, Jamie.

ACKNOWLEDGEMENTS

I have many people that have helped me along the way to this point in my life. I can only hope to repay that debt by treating others with the great generosity that has been bestowed on me all these years.

I would like to thank my advisor, Tony Yezzi. He is a selfless, patient, brillant yet humble mentor. I would not have come this far without his guidance and infinite patience.

Many thanks go to Professor Allen Tannenbaum whose humor, wit and, intelligence never seem to fail.

I thank the many people who helped me edit this dissertation: my parents Dr. Thomas and Louise Jackson, Ms. Linda Lambert, Professors Allen Tannenbaum and Aaron Lanterman.

I would like to thank the members of my committee: Professors Allen Tannenbaum, Greg Turk, Aaron Lanterman, and Patricio Vela.

I have enjoyed working with so many collegues in my lab: Ganesh Sundaramoorthi, Chris Alvino, Siddharth Manay, Gozde Unal, Carol Li, and Miguel Lopez.

I would like to thank my friends over the years: Juan Barahona, Miguel Salgado, etc...

I could not have made it without the support of my friends at the Catholic Center at Georgia Tech, Fr. Mario DiLella and everyone else.

I would especially like to thank my family:

- My Father for inspiring me to follow in his footsteps and get a doctoral degree.
- My Mother for her support, love, and guidance.

- My brother, Jonathan, for inspiring me to follow Electrical Engineering and Computer Science as a major.
- My brother, Joshua, for inspiring me to attend Tulane.
- My sister, Jamie, for the joy she brings into everyone's life.

I am truly blessed...

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

When modeling a dynamic scene, i.e., an image sequence which contains one or more moving objects, a number of factors must be taken into account. The background can appear to deform and move because of camera motion and distortion effects from the lens, and the background may also change on its own. Furthermore, foreground objects may occlude other foreground objects and the background as they deform and move in the scene. The appearance of the background and foreground objects may be altered by viewpoint changes of the camera and shadowing effects. Clearly, faithful modeling of such a scene appears to be quite challenging; thus, methods approach modeling by taking a restricted portion of the scene or giving a simplified dynamical scene description.

This dissertation gives a general model for the estimation of shape (image segmentation), appearance, pose (image registration), and movement (tracking). The model can infer parameters for multiple objects in a dynamically changing scene. There are a number of real-world applications. In particular, in visual tracking, moving the camera to keep objects of interest in the field of view may cause the background to move. The objects can move and deform in three dimensions, but they must be captured in two-dimensional images.

Each component of the image is represented by a separate layer: one for the background and a layer for each foreground object. Each layer has three components: a contour that bounds the region of the layer, a smooth function that represents the object's appearance, and a transformation that maps that layer into an image. The segmentation for each layer is a contour (embedded as the zero level set of a

distance function) that is the average shape of the object computed from multiple images. The smooth function associated with a layer approximates the image data inside the contour, after the contour has been mapped into the image by a similarity transformation (rigid component) plus a vector field (non-rigid component). A practical application of having this model is that one can fix the size of a layer and then construct priors on both shape and appearance for that layer. These priors are constructed using principal components analysis (PCA), which reduces the dimensionality of the image-approximating smooth function and the vector field (non-rigid registration) and allows for more accurate modeling of an object for that layer.

CHAPTER I

INTRODUCTION

1.1 Motivation

When modeling a dynamic scene, i.e., an image sequence containing one or more moving objects, a number of factors must be taken into account. The background can appear to deform and move because of camera motion and distortion effects from the lens, and the background may also change. Foreground objects may occlude other foreground objects and the background as they deform and move in the scene. The appearance of the background and foreground objects may be altered by viewpoint changes of the camera and shadows from the scene. Clearly, faithful modeling of such a scene appears to be quite challenging; thus, many methods approach modeling by taking a restricted portion of the scene or giving a simplified dynamical scene description. In the next section, the literature corresponding to these methods is reviewed and discussed.

1.2 Origin and History of the Problem

This dissertation covers a number of open research areas in computer vision: Grenander's pattern theory [72], active shape models and active appearance models, image inpainting and diffusion (isotrophic and anisotrophic), layered representations, segmentation, registration, and tracking with all of its methods and error correction.

This doctoral research is strongly connected to pattern theory and matching, which includes Yuille's deformable templates [207]. Active shape models and active appearance models by Cootes [49] also fit into this scheme.

In addition to matching data, the model given in this dissertation fills in data

much like image inpainting. Image inpainting is the restoration of parts of an image where data is missing [16]. The methods proposed here use a diffusion term in the energy that is minimized to find the smooth function that approximates the image data. An alternate way to enhance an image is through anisotrophic diffusion [145]. Speed is always a factor when numerically solving for partial differential equations (PDEs). Fast methods for non-linear diffusion include additive operator splitting [190], conjugate gradient descent, and multigrid methods [31].

1.2.1 Layered Representations

The model used in this dissertation represents each object by its own layer. This layer is then mapped into the image by a rigid transformation composed with a non-rigid transformation to model the layer's corresponding object in the image. There has been quite a large literature devoted to this subject. Wang and Adelson proposed a layered model with an affine transform for moving the layers around [187]. They proposed, but did not implement a non-rigid component. Ayer and Sawhney [9] use maximum likelihood, expectation maximization (EM), and minimum description length to find layers in images. Hsu, Anandan, and Peleg use layers to improve optical flow computation [79]. Zhang, Paragios, and Metaxas [210] find the layers in an image by using optical flow, clustering, and level sets. Frey, Jojic, and Kannan [68] use layers with rigid and non-rigid registration and segmentation in a way similar to the method in this dissertation, but they use the EM algorithm, whereas this dissertation uses a variational method.

1.2.2 Segmentation using Active Contours

Segmentation is the process of extracting region(s) of interest from an image. It is often based on the minimization of some energy according to some corresponding metric. The method described in this dissertation can reduce to the solution of segmentation using active contour models. Active contour models (deformable curves,

which may be used to segment an object in the image) were first introduced in the seminal paper of Kass, Witkin, and Terzopolous [105]. Cohen added a balloon term to active contour models [46]. The implementation of these methods used parametrized curves. Some other implementations use Osher and Sethian's geometric model for curves, so-called "level sets" [134]. Kichenessamy *et al.* and [108] and Caselles *et al.* [184] proposed capturing objects in an image via level-set based active contours that are attracted to strong image gradients (that indicate image edges). Mumford and Shah [130] employ a piecewise smooth function to approximate image data and optimally segment a region of interest from an image. Chan and Vese [38] simplified this model so the image data is approximated by one constant inside the region and a different constant outside the region (usually the image means). Yezzi *et al.* [205, 204] extended the work of Mumford and Shah [130], and Chan and Vese [38], with piecewise constant and piecewise smooth functions. In addition to geometric-based level sets and parametrically-based snakes, active polygons have been used for segmentation as well by Unal *et al.* [182].

Several techniques may be employed to speed up segmentation algorithms in this context. Methods employing level sets generally are computed using a narrowband structure, which computes forces in a small band around the zero level-set which avoids computing the update for the level set over the domain of the whole image. Kuhne [113] and Weickert [188] use additive operator splitting for speed. Kimmel uses multigrid methods to solve for active contours [107] for speed and possibly a more global minimum.

1.2.3 Registration

Image registration involves finding a transformation that matches the pose of an object in an image to a similar object in another image. The transformation can be rigid, affine, or non-rigid. The method in this thesis maps a layer into an image

via a transformation to match the pose of an object that the layer is modeling. Registration can be done with the whole image or a region of interest. Registration is usually performed using noticeable landmarks, general image information, or a given segmentation. There is a huge literature devoted to image registration; only some of the references most relevant to the work of this dissertation will be mentioned. The difference between using purely image information and shape information is discussed in the work of Cachier *et al.* [35]. In [199, 197], the authors treat the problem of simultaneously finding the registration and segmentation of an object in multiple images, with the individual solutions of the registration and segmentation feeding information to each other. Mutual information is a measure that may be employed for registration [186] as well. A number of registration techniques using mutual information can be found in Pluim *et al.* [148]. A number of non-rigid image-based registration techniques are based on optical flow [78, 131, 117]. A comparison optical flow techniques may be found in the survey by Barron [13]. Alvarez *et al.* explore changing the regularization term of the non-rigid registration to obtain large optical flow displacements [6]. A non-rigid registration technique based on mass transport was proposed in [75]. Clarenz uses multigrid methods for fast [44] non-rigid registration [43]. Huang *et al.* [82] employ spline-based free form deformations to constrain non-rigid registration, while Frey *et al.* [68] use the EM algorithm for rigid and non-rigid registration.

1.2.4 Tracking

As with the preceding subjects, there is a large literature on tracking. A number of researchers use the Kalman filter in conjunction with some parameterized snake model [173, 146, 147] since these are amenable to state space methods. Paragios *et al.* [137, 139] have performed tracking by using geodesic active contours and regions. Yezzi and Soatto devise a method for tracking registrations [197] that is implemented

using an observer [88] that estimates the infinite-dimensional level-set representation of the curve. Both Niethammer *et al.* [132] and Rathi *et al.* [150] have methods for tracking curves and statistical methods for dealing with occlusions. Blake and Isard used particle filtering for tracking parameterized curves [85, 22]. Data association techniques [12] must be in tracking as in Hager [149] and Peterfreund [146].

1.3 Contributions

The main contribution of this dissertation builds upon and unites earlier results. This dissertation proposes a model for the estimation of shape (image segmentation), appearance, pose (image registration), and movement (tracking). The concept of *layered deformotion with radiance* models these parameters for multiple objects in a dynamic scene. As mentioned above, in visual tracking, moving the camera to keep objects in the field of view will cause the background of the images to move. The objects of interest can move and deform in three dimensions, but they must be captured in two-dimensional images. Each component of the image is represented by a separate layer for which there is a layer for the background and a layer for each foreground object. The segmentation for each layer is a contour (embedded as the zero level set of a distance function) that is the average shape of the object computed from multiple images. The smooth function associated with a layer approximates the image data inside the contour, after the contour has been mapped into the image by a similarity transformation (rigid component) plus a vector field (non-rigid component). A practical application of having this model is that one can fix the size of a layer and then construct priors on both shape and appearance for that layer. These priors are constructed using principal components analysis (PCA), which reduces the dimensionality of the smooth function that approximates the image data and the vector field (non-rigid registration) while allowing for more accurate modeling of an object for that layer.

The segmentation and registration are found by minimizing a quadratic energy functional that compares the image intensities of an object against some smooth function that represents that object. This smooth function can depend on the image intensities inside the curves capturing the objects. In this dissertation, a combination of learned smooth functions that are the principal components of a training set model that particular object's image intensities is employed. The smooth function is computed by using gradient descent during the training phase. A forward Euler scheme is then used to solve the resulting partial differential equation (PDE). The vector field and the segmenting curve are also computed via the Calculus of Variations.
In summary, the key contributions of the doctoral work described in this dissertation are:

- A segmentation method is presented for objects with regions of faint data (e.g., dendrites and their spines with their necks and heads). It incorporates a topology preservation method to include the areas with little or no data and to keep the object as a single object. It uses multiscale diffusion to coarsely segment the object and then refine the segmentation by diffusing less and less.

- A novel method for tracking objects in a scene is shown that incorporates the segmenting curve as the state along with the parameters of a rigid registration that map that curve into each of the images. An observer is used to show success in tracking under severe occlusions.

- The main model of this dissertation is presented for the estimation of shape (image segmentation), appearance, pose (image registration), and movement (tracking) in a scene. A separate layer is used to represent each component of the image. As above, each layer has three components: a contour that bounds the layer region, a smooth function that represents the appearance, and a transformation that maps that layer into an image. It will be shown how this

method abstracts many existing methods in computer vision such as L1 and L2 optical flow, motion segmentation, and image inpainting.

- Finally, the results of layered deformotion (specifically the smooth function that approximates the image data and then the non-rigid registration) that are run on a database of images are used to build a prior on both shape and appearance to better model new images of the same class. Principal components analysis (PCA) is used on the smooth image-approximating function and the non-rigid registration. This information is put into the layered deformotion framework for faster and more robust modeling.

CHAPTER II

MULTISCALE DIFFUSION WITH MUMFORD-SHAH

2.1 Overview

This chapter presents a segmentation algorithm that uses multiscale diffusion with the Mumford-Shah model. The Mumford-Shah model is a generative model that employs smooth functions to approximate the image data inside and outside of the a surface that segments an object in the image. Those image-approximating functions are formed with a trade-off between smoothing and data fidelity in the energy functional. The smooth functions are solved for by using a partial differential equation to minimize an energy functional. Here, a scale-space approach is used that employs a good deal of diffusion as its coarse scale space to initially solve for the smooth functions and get a coarse-scale segmentation. Gradually, the diffusion is reduced to yield a fine-scale segmentation that acquires more features of the foreground object. The algorithm continually moves to a particular diffusion level, rather than just using a set diffusion coefficient with the Mumford-Shah model. Each time the smoothing is decreased, the data fidelity term increases and the surface is moved to a steady state.

This method is useful in segmenting biomedical images acquired using high-resolution confocal fluorescence microscopy. The method is tested on images of individual dendrites of neurons in a rat visual cortex. These dendrites are studded with dendritic spines, which have small heads and faint necks. The coarse-scale segmentation acquires the dendrite and the brighter spine heads, while avoiding noise. Backing off the diffusion coefficient on the image-approximating function to a medium scale fills in more of the structure of the dendrite, which obtains some of the brighter spine necks. The finest scale fills in the small and detailed features of the spines that are

missed in the initial segmentation. Because of the thin, faint structure of the spine necks, a topology preservation method by Han *et al.* [76] for the surface is incorporated into the level set framework, which aids in segmentation and keeps a simple topology.

2.2 Background on Segmentation

Global segmentation algorithms have the benefit of being able to extract an object and its prominent features from an image or image volume. They have this capability because they segment an image by using image properties such as the average pixel intensity of a region or differing textures of regions. Some of these methods are detailed in [205, 38, 140, 151, 211], and the references therein. While a global perspective is better able to avoid the noise than more edge-based detectors, it may lose the fine-scale features of the object in capturing a coarse estimate of the object. It would be useful to have a method for including some of these finer scale features after this coarse segmentation has been done. A method of this type was necessary to solve the problem of segmenting a topographically complex biological structure from a three-dimensional image volume.

The structures in question are the dendrites of pyramidal neurons in a rat visual cortex. These dendrites are studded with individual tiny branchlets called *spines*. The spines are sites of synaptic contact between neurons, and their three-dimensional morphology is thought to be a marker of the functional state of individual synapses. The fine structure of spines has been extensively investigated at the electron-microscope level—they are known to be bulbous in shape and always connected to the dendrite by thin necks (with diameter on the order of 0.1 micron) [133].

Three-dimensional image volumes of spiny dendrites were obtained as follows. Pyramidal neurons in fixed tissue slices of a rat visual cortex were intracellularly injected with the fluorescent dye Alexa-488 (Molecular Probes Inc., Eugene, OR;

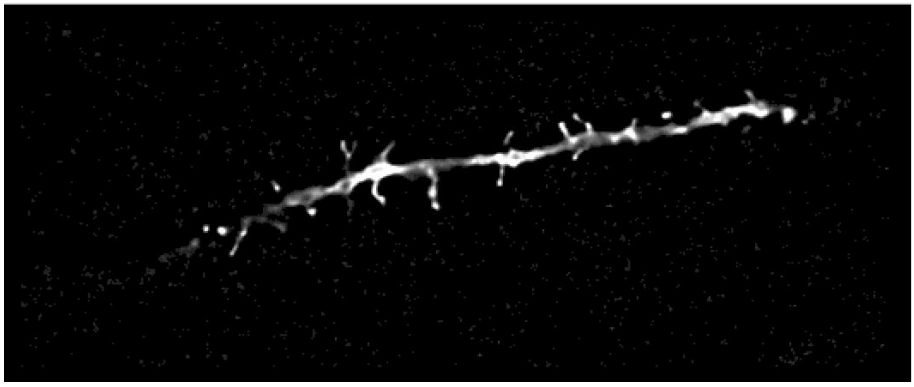

Figure 1: *2D image plane from the middle of a three-dimensional volume, showing the dendrite with spines branching off.*

emission peak = 517 nm). Individual dendritic segments were imaged in stacks of two-dimensional images by using an Olympus fluorview confocal microscope, at zoom factor 8, with a 63x NA 1.2 water-immersion lens. The voxel size of these images was 0.09 x 0.09 x 0.15 microns (slightly above the diffraction limit of this imaging system). Three-dimensional images were preprocessed by using simple operations to improve contrast and reduce noise. Images were then deconvolved by using an adaptive blind deconvolution algorithm (Autoquant Imaging, Watervliet, NY).

In these images, the dendrite is more brightly fluorescent than the spines, due to the greater volume of fluorescent dye it contains. The spine necks, in particular, can be quite faint both because of their small volume and their size being at the limit of resolution of the confocal microscope. Some of the spine heads are dim as well. It is apparent that some of the spine heads are dim when ones views a full two-dimensional slice of the three-dimensional images as in Figure 1 and a close up of a section of the dendrite and its spines as in Figure 2.

The region-based methods only capture the dendrite and some of the spine heads. The first step to solve this problem is to set a smoothing parameter in the Mumford-Shah segmentation method so that it becomes a region-based algorithm that gives a coarse segmentation of the dendrite. This smoothing term is gradually reduced to capture some fine-scale features. This stepping down of the diffusion term gradually

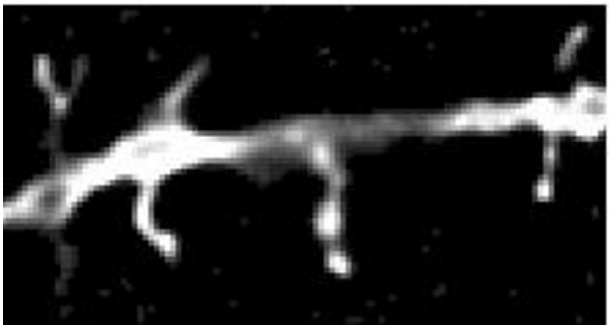

Figure 2: *Closeup of Figure 1, showing the fine structure of spines. Pixel size = 0.09 x 0.09 microns.*

yields a correct segmentation of the spine heads and the necks that connect them to the dendrite.

2.3 The Mumford-Shah Model

This section presents the variational formulation of the main segmentation algorithm (a multiscale version of Mumford-Shah). This algorithm was implemented in a level set formulation according to Osher and Sethian [134]. Other level set implementations of Mumford-Shah are in [205, 37], and the model itself is in the original paper by Mumford and Shah [130]. The level set is evolved by using a PDE that minimizes a given energy functional. More about level set theory can be found in the book by Sethian [159].

To implement the Mumford and Shah model, smooth estimates of the foreground and of the background are required to find the piecewise smooth estimate of the image data together with the surface of discontinuity separating the two estimates. Based on smooth estimates, the three-dimensional surface is evolved to minimize the energy functional:

$$\begin{aligned} E &= \alpha \int_R (I-f)^2 dV + \alpha \int_{R^c} (I-h)^2 dV \text{ (data fidelity)} \\ &+ \beta \int_R |\nabla f|^2 dV + \beta \int_{R^c} |\nabla h|^2 dV \text{ (smoothness)} \\ &+ \gamma \int_S d\sigma \text{ (surface area penalty)}, \end{aligned} \tag{1}$$

where I is the image volume, f is the smooth estimate of the image in the foreground R, h is the smooth estimate of the image in the background R^c, and S is the segmenting surface.

The first two terms in the energy functional (1) are data fidelity terms that ensure that the smooth estimates of the foreground and background match the image data as closely as possible. The next two terms keep the norm-squared gradients of the smooth estimates f and h as small as possible, which results in a smoother f and h. The last term of the energy functional is used to penalize surface area. The parameters $\alpha, \beta, \gamma \in \Re$ are weights that increase data fidelity, or smoothness, or penalization of surface area. The surface S is evolved according to the flow

$$S_t = -\alpha((I-h)^2 - (I-f)^2)\vec{N} + \beta(|\nabla h|^2 - |\nabla f|^2)\vec{N} + \gamma\kappa\vec{N} \tag{2}$$

where $\vec{N}$ is the inward normal of the surface S [176]. With each evolution time-step of S, the new smooth estimates f and h are computed by using the same energy functional as above (1). Minimization of that energy is performed with respect to f when evolving the smooth estimate f. Using the Calculus of Variations, the first variation obtains the Euler-Lagrange equations necessary to evolve the smooth function to a steady state based on the the energy functional. The resulting equation to evolve the smooth function f is

$$f_t = 2(\alpha(I-f) + \beta\Delta f), \tag{3}$$

where Δf is the Laplacian of f:

$$\Delta f = f_{xx} + f_{yy} + f_{zz}. \tag{4}$$

Evolving h is similar.

A piecewise constant version of (1) is given in [38]:

$$E = \alpha \int_R (I-u)^2 dV + \alpha \int_{R^c} (I-v)^2 dV + \gamma \int_S d\sigma. \tag{5}$$

where u and v are the means inside and outside the surface, respectively. The evolution of the the surface S is then

$$S_t = -\alpha(u-v)(I-u+I-v)\vec{N} + \gamma\kappa\vec{N}. \tag{6}$$

The Chan-Vese flow can also be viewed as the (β=∞) case (total smoothing) of the Mumford-Shah flow.

2.4 Multiscale Diffusion with Mumford-Shah

In the algorithm presented in this chapter, the initial coarse Mumford-Shah segmentation uses $\beta = \infty$, which is equivalent to the Chan-Vese piecewise constant model. The Mumford-Shah flow is evolved to steady state while gradually decreasing the smoothing parameter.

The $\beta = \infty$ case corresponds to a coarse scale space, which will only segment prominent features of the image. The update of the surface can be expressed as

$$S_t = -2\alpha(u-v)(I - \frac{u+v}{2})\vec{N} + \gamma\kappa\vec{N}. \tag{7}$$

If the surface is initialized so that it is outside of the object that one wants to segment, then the sign of the term $-2\alpha(u-v)$ should not change while the surface is evolving. The term $I - \frac{u+v}{2}$ shows that the flow will move the surface according to u (the mean of the image data inside the surface) and v (the mean of the image data outside the surface) so that the energy

$$E = \alpha \int_R (I-u)^2 dV + \int_{R^c} (I-v)^2 dV \tag{8}$$

is as small as possible. With each iteration, the means are computed and the surface moves past a pixel in I if it is less than $\frac{u+v}{2}$. This is the case if one ignores the surface area penalty. The value $\frac{u+v}{2}$ in this case can be looked at as a threshold that gets larger as the surface segments a bright object. This flow gives a segmentation of all of the prominent features of the object. The problem with this segmentation is the

the single value $\frac{u+v}{2}$ that is used to move the surface at all points in the image. This value tends to skip over fine detail that might be fainter than most of the rest of the object. In the case of dendrites, the main dendrite and the head of the spines are segmented quite well, but the dimmer spine necks are completely skipped over.

To correct this problem, one needs the Mumford-Shah flow with $\beta \neq \infty$ which uses a value $\frac{f+h}{2}$ to decide whether or not to pass by a pixel. Since f and h are smooth functions, there is a more adaptive threshold that passes by pixels depending on a value that is more local to the pixel, since f or h at each pixel is smoothed by its neighboring pixels. This method is preferred over a global smoothing ($\beta = \infty$), which results in $f = u$ and $h = v$. The more local version allows Mumford-Shah to capture some of the finer details as well. The premise of the algorithm is to keep backing off the smoothing to acquire more details of the object from a coarse initial estimate. This gradual acquiring of features in a multi-step fashion allows the flow to accurately capture more detail than a Mumford-Shah flow with a set diffusion coefficient. The fixed Mumford-Shah flow does not capture such details as well as this multi-step version, because it has no good coarse segmentation to build upon, as can be seen in Figure 3. An assumption that was made in the segmentation of dendrites is that the background is constant (fairly close to zero), which turns out to be true for all the dendrite data used in this research. This allows one to use v or zero as the estimate for the background, which speeds up the process since it is not necessary to use a PDE to find the smooth function h each time the surface needs to be evolved.

2.5 Topology Preservation with Mumford-Shah

The segmentation of an object should be as realistic as possible. Dendrites have no holes of any kind, so the segmentation of a dendrite should be topologically equivalent to a sphere. To accomplish this, one needs a flow that preserves the topology of

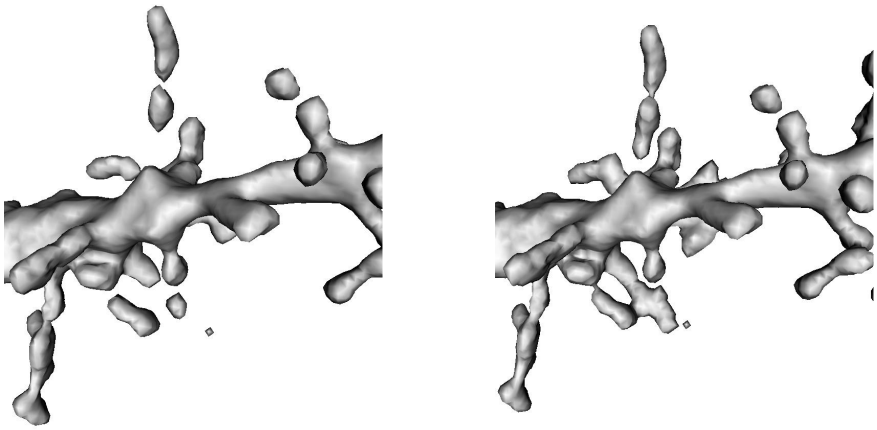

Figure 3: *Set Mumford-Shah versus multiscale Mumford-Shah.*

the surface that represents the segmentation. This flow will also preserve fine-scale features (i.e. the necks of the dendritic spines). The topology-preserving method was originally developed by Han *et al.* [76]. If topology preservation is not used, the surface will pinch off at the necks and will only segment the main dendrite and the heads of the spines. Figure 4 shows a comparison of using topology preservation versus not using topology preservation.

A level set function uses values below zero to denote the inside of the surface (the zero level set) and values above zero to denote pixels that are outside the surface. When a value of the level set ϕ wants to change sign, i.e. a pixel wants to change from foreground to the background, etc., that change may cause a change in topology. To stop a change from occurring, one must check to see when a point in the level set will change sign. If this change will cause a break in topology, the value of the level set is set to be some small number ϵ that has the same sign as before. For more in-depth information on topology preservation one should see Appendix A.

Topology preservation helps at every point in the evolution of the surface. The initial thresholding flow is equivalent to a Chan-Vese flow. The Chan-Vese flow has to preserve topology; otherwise, in the case of dendrites, the necks would get pinched off. It is possible to get these necks returned without doing topology preservation

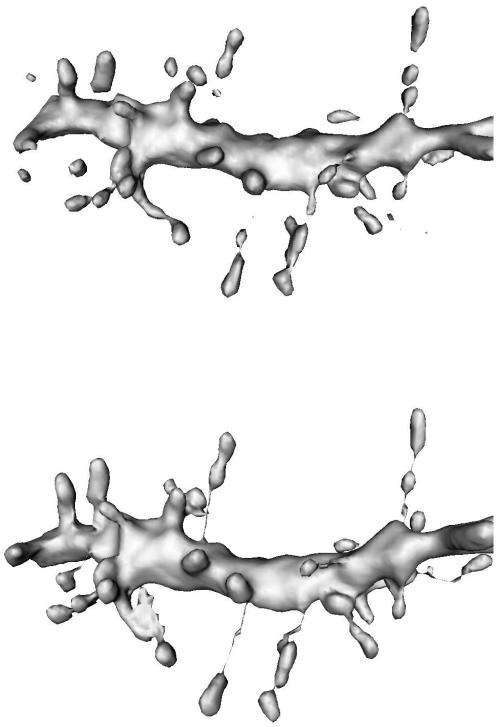

Figure 4: *Mumford-Shah* ($\beta = \infty$) *case: no topology preservation versus topology preservation.*

and just running the multi-step Mumford-Shah. The advantage of having topology preservation is that there is a piece of surface that is already connecting the spine head and the main dendrite where the neck should be. This surface connection makes it easier for the multi-step Mumford-Shah to expand out over that neck. In the case of a missing neck or extremely faint data, the multiscale Mumford-Shah will not fill in the neck and have it fully connect the spine head to the dendrite. Figure 5 shows a comparison of using multiscale Mumford-Shah without and with topology preservation is shown.

2.6 Conclusion of Dendrite Segmentation

This chapter has shown a method that segments fine-scale features of a biological object. This approach of a multiscale Mumford-Shah algorithm is good for first capturing coarse-scale features and then fine-scale features. The topology preservation

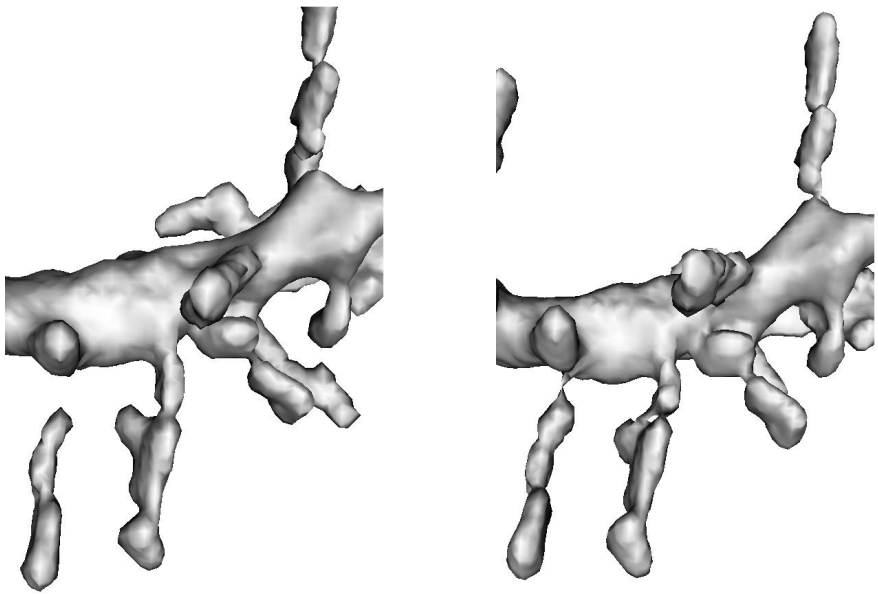

Figure 5: *Connection example: multiscale Mumford-Shah without and with topology preservation.*

allows for a more realistic segmentation with no breaks in topology. With prior knowledge of dendritic spines and their topology, one can require that spine heads remain connected to the dendrite by a neck, even when there is no data supporting a neck. Topology preservation also improves segmentation of spine necks in cases where the data for the neck exists but is faint. This improvement is evident in Figure 6 where the progression of multiscale Mumford-Shah is shown with and without topology preservation. The multiscale Mumford-Shah with topology preservation captures the dendrite quite well.

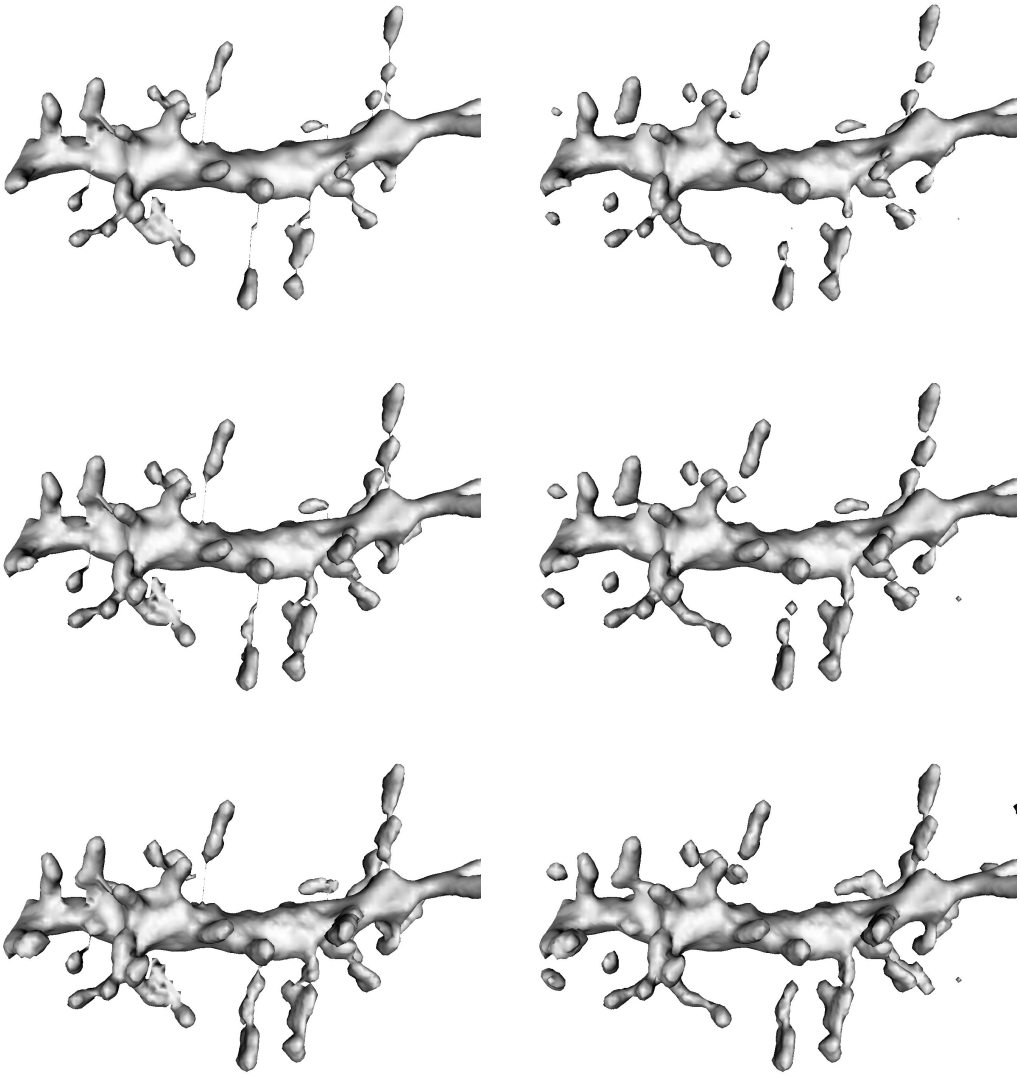

Figure 6: *Progression of multiscale Mumford-Shah with (left column) and without (right column) topology preservation.*

CHAPTER III

TRACKING USING DEFORMOTION

3.1 Overview

This chapter describes a nonlinear model for tracking a slowly deforming and moving contour, despite significant occlusions. The contour is represented implicitly, and its motion is described by the action of a finite-dimensional group. Both the implicit representation of the contour (its shape) and its motion are estimated. The contribution of this chapter consists in defining a generative model that is not subject to arbitrary re-parameterization, not subject to choice of (non-unique) key points or control points and enforces a dynamical model of motion when it is available. The approach may incorporate simple phenomenological models, such as low acceleration or low jerk.

3.2 Background on Tracking

This chapter considers the problem of tracking moving objects in a sequence of images. An "object" is a region of the image that has a distinct photometric signature, i.e., something that distinguishes it from the rest of the image, or "background." For example, it could have quasi-homogeneous intensity, or some other statistic that is uniform or almost uniform within the object, but distinct from the rest of the image. In particular, it is helpful to be able to to track the object despite its being invisible or partly invisible at certain instants of time. The object should be tracked despite changes in the geometry, and possibly the topology, of the region that determines it.

The issue of deformation has received significant attention in the literature, which

is reviewed briefly in the next subsection. In particular, Yezzi and Soatto [197] characterize the tracking of a moving object, where motion can be defined by a finite-dimensional group (e.g., affine), through the introduction of a generative model of the so-called "average shape," from which each measurement is obtained with minimum deformations, measured with respect to a chosen energy functional. While that work hinted at the problem of extending the framework to the case when the underlying shape average is changing over time, it did not offer a technical solution for tracking with an explicit dynamical model.

To track regions through occlusions, a motion model is necessary to predict, or extrapolate, the state of the object when one lacks all the measurements. This issue of learning motion dynamics has also gathered considerable attention in the past, and indeed some of the most popular particle filtering techniques were developed in the context of contour tracking [22]. In this dissertation, however, objects are considered as regions bounded by closed planar contours, represented implicitly. These are infinite-dimensional objects, and there is no existing filtering technique suitable for such infinite-dimensional state-spaces. Therefore, this chapter represents a small step in a significantly novel and challenging direction, as is described in the next subsection. Although a rigorous solution to this problem is elusive, approximate filtering can be performed in a way that results in simple, robust and efficient algorithms that are validated experimentally in Section 4.6 on sequences of images with severe occlusions.

3.2.1 Relation to the State of the Art

Contour tracking has been an active research area in vision for many years. The book of Blake and Isard contains a snapshot of the state of the art as of 1998 [22]. What makes the problem different from a standard tracking problem, as studied in signal processing and control theory, is the fact that the representation of the state of the

model and of the measurement map is non-trivial. In traditional tracking the "target" is usually a point or a collection of points in a vector space. In particular, a common approach in contour tracking is to represent the contour through a finite-dimensional representation. Different representations include various types of splines or "snakes" (see [22, 106] and references therein) and various discretizations of the contour; for instance, by using polygonal approximations [182, 181].

In all these finite-dimensional approximations, a dynamical model is introduced by modeling the *parameters* (e.g. the nodal points, or the control points, of the representation) as the state of a dynamical model, typically assumed linear (e.g. an autoregressive moving-average model). The difficulty with this approach is that each contour is represented *not* by a set of parameters, but by an entire equivalence class of parameters obtained by moving the control/nodal points. There are infinitely many choices of control points that result in substantially the same measured contour. Therefore, many have resorted to additional constraints; for instance, equi-spaced polygonal vertices, a fixed number of equi-distant control points, etc. Additional techniques rely on describing regions by using "blobs" or other objects with pre-specified shape, or by collections of spatial configurations of blobs. Using blobs is common for the case of cars and people (see, for instance, [29, 20] and references therein); such techniques have proven successful even in the case of a severe occlusion [149].

A substantially different approach is taken when the contour is represented in the continuum. For instance, "deformable templates," pioneered by Grenander [72], do not rely on "features" or "landmarks"; rather, images are directly deformed by a (possibly infinite-dimensional) group action and compared for the best match in an "image-based" approach [207, 10, 206, 128, 110, 156, 111, 109, 38, 116]. A common model is to implicitly represent the contour, as the zero level set of a function (e.g. a signed distance function) that evolves in time. Geodesic active contours [108, 184]

have been successfully used for tracking; one application being tracking cardiac motion in ultrasound imaging [137]. In most current approaches, however, "time" only indicates the index of the iterative procedure used to estimate the contour. Most motion models are essentially assuming that the position of the object at time $t+1$ is close to that at time t; and therefore, the best estimate of the contour at time t can be used to initialize the same procedure at the next instant [129].

This chapter considers higher-order motion models as might arise from inertia and other constraints on acceleration. The idea is to set up a framework where a detailed motion model can be used if one is available. If one is not available, statistical or phenomenological motion models can be used. In particular, one may want to enforce regularity by imposing that the velocity, acceleration, or jerk (the derivative of acceleration) is small.

Ideally, one would like to derive an optimal framework to enforce higher-order motion models. This action would entail estimating the conditional density of the state (motion and shape of the deforming contour) given the measurements up to time t (noisy/deformed measurements of the contour, possibly with significant missing pieces). For linear dynamical models driven by additive, white, zero-mean Gaussian noise, this framework is appropriate. However, this framework does not seem feasible for a state that is infinite-dimensional (the shape of the deforming contour), has non-trivial geometry (the group structure), highly non-linear measurement equations (due to occlusions), and uncertainty that is functional, rather than additive (the diffeomorphic model of the contour deformation). Therefore, one can only resort to approximate filtering techniques, with no available analytical statements about their performance. Filtering for non-linear finite-dimensional models has received a lot of attention since the age of Wiener in the mid forties, and has culminated in several viable (although generally not optimal) techniques, such as the Extended Kalman Filter [91], the multimodal sum-of-Gaussian filter [3], particle filtering [22], various forms

of multimodal, multi-target tracking based on interacting multiple models (see [12] and references therein) and various numerical approximations of the Mortensen-Zakai equation. However, there has been little work on filtering for infinite-dimensional state spaces. Blake and Brockett [21] addressed the problem of estimating a moving curve (represented as the graph of a function) despite missing segments of the curve. In this thesis, the problem is more difficult because one cannot rely on the graph structure; and furthermore, the solution is entirely different from that suggested in Blake and Brockett's paper [21].

3.3 Formalization of the Problem

At any instant of time $t \in \mathbb{R}$, let $\mu(t) : \mathbb{S}^1 \to \mathbb{R}^2$ be a closed planar contour, $g(t) \in G$ be a finite-dimensional group action (e.g. the Euclidean group $G = SE(2)$ or the affine group $G = A(2)$), and $h_t : \mathbb{R}^2 \to \mathbb{R}^2$ be a diffeomorphism that can change over time (hence the subscript t). The measurement of a closed planar contour is $y(t) : \mathbb{S}^1 \to \mathbb{R}^2$, a local deformation h_t of the static contour μ moving under the action of g. Therefore, a generative model for the data $y(t)$ can be formally written as:

$$\begin{cases} \dot{\mu}(t) = 0 \\ \dot{g}(t) = \widehat{v} g(t) \\ \dot{v}(t) = \alpha(t) \\ y(t) = h_t(g(t)\mu(t)). \end{cases} \tag{9}$$

In this model, the first equation embodies the assumption that the average shape is constant. If this is not the case, but still it varies slowly relative to the intrinsic dynamics of $y(t)$, one could write formally that $\dot{\mu}(t) = w(t)$. If there are no assumptions made on how the average shape evolves, the tracking problem cannot be meaningfully addressed [197]. The second equation is a deterministic integrator that states that pose is the integral of velocity, and the third equation states that velocity is the

integral of acceleration, which is unknown. It could be assumed that α is an unknown constant or that it varies slowly relative to the dynamics of $y(t)$. Finally, the last equation states that the measurements are a perturbation of the average shape in the moving frame. The goal is to infer μ, g and v from measurements of y. In particular, the interest is in the estimate that results in the "smallest" possible perturbation $h_t(\cdot)$. This model is a formal notation; to design, implement and analyze inference algorithms, one must specify (i) a representation for μ, (ii) a local coordinate system for g, and (iii) a discrepancy measure between the data $y(t)$ and the model of the data, $h_t(g(t)\mu(t))$. The second issue is straightforward since canonical coordinate charts for matrix Lie groups are easy to derive and compute by using the exponential map [118]. The first and the third issues are more complex and highly interconnected.

Consider a parametric representation of $\mu(t)$, e.g., $s \mapsto \tilde{\mu}(t, s)$. The measured contour $y(t)$ can also be parameterized via $l \mapsto \tilde{y}(t, l)$. Unfortunately, the *correspondence* between s and l is not known; therefore, the measurement equation relies on an estimate of the reparametrization $l \mapsto s = \rho(l)$, or on a canonical representative of the equivalence class. This significantly complicates the model, since one now has $y(t) = \tilde{y}(t, l) = h(g(t)\tilde{\mu}(t, \rho(l)))$ and no constraints on ρ other than its being a smooth bijection. Therefore, one may choose to represent μ *implicitly* as a set $\mu(t) = \{x \in \mathbb{R}^2 \mid \chi_\mu(x, t) \leq 0\}$, where χ denotes the characteristic function of the set μ. While representing μ *implicitly* as a set causes no problem for the basic model where $\dot{\mu} = 0$; in the case of "slowly varying" average shape, one has to define what $w(t) \doteq \dot{\mu}(t)$ means. Furthermore, the lack of a correspondence $s \leftrightarrow l$, forces a specification how one computes a discrepancy between $y(t)$ and $h_t(g(t)\mu(t))$. Define $w(t)$ as the quantity specified by the equation $\mu(t + dt) = \mu(t) \oplus w(t)$, where $\oplus$ denotes a composition operation in the set $\mu(t)$, e.g., $w(t)$ can be the set-symmetric difference between $\mu(t + dt)$ and $\mu(t)$.

3.4 Filtering Deforming Shapes

With this formalism, one can postulate the structure of the state estimator. The starting point is $t = 0$ with an initial point estimate, $\{\hat{\mu}(0), \hat{g}(0), \hat{v}(0)\}$. Since the global reference is arbitrary, one typically chooses $\hat{g}(0) = Id$, the group identity. At time t, in absence of any new measurement, the best estimate of the state at $t + \delta$ is obtained by integrating the state equation between t and $t + \delta$. The unit of time is δ; integrating the basic model for constant μ and constant velocity is straightforward, since $\hat{\mu}(t + 1) = \hat{\mu}(t), \hat{v}(t + 1) = \hat{v}(t)$ and $\hat{g}(t + 1) = \exp(\widehat{\hat{v}}(t))\hat{g}(t)$. Here " $\widehat{\ }$ " is the operator onto the Lie algebra, and the superscript "ˆ" denoting an estimate is omitted from $v(t)$ for ease of notation.

Assuming that a new measurement $y(t+1)$ becomes available, the interest is in updating the prediction in a way that guarantees that the estimate of the state will converge, asymptotically, to the true state. While one can derive the optimal estimator directly in the case of a linear finite-dimensional model, there is no finite-dimensional optimal estimator in general. Therefore, one needs to postulate a structure of the estimator, in the form of a generic local observer, and then choose parameters that guarantee error stability.

Since the deterministic integrator $\dot{g} = \widehat{v}g$ does not impose any model constraint (other than adherence to the group action G), that equation carries no uncertainty; if v was known exactly and g was known exactly, then $\dot{g}$ would be given *exactly* by $\widehat{v}g$. Therefore, that equation carries no model error, and the filter can be saturated along the corresponding components. In practice, one writes the equation in local coordinates Ω, defined by $g \doteq e^{\widehat{\Omega}}$ and approximates the equation to first order as $\dot{\Omega} = (I + \widehat{\Omega})v$. Consequently, the measurement equation becomes, $y = h(e^{\widehat{\Omega}}\mu)$ (neglecting the time indices).

The goal of the update step is to reduce the "uncertainty," i.e. the discrepancy between the model and the measurements. The uncertainty is the diffeomorphism

h_t, not a usual additive noise; at each step, one has to solve a local optimization to compute the best update for the state. In particular, a local update is considered that is based on an incremental step in the direction of the gradient of a cost functional that measures the "energy" of the diffeomorphism h_t, subject to the model (9). Specifically, at time t, a causal window of length $k \geq 2$, and the energy

$$\phi(v_1, \ldots, v_k, \Omega, \mu) \doteq \sum_{\tau=t-k+1}^{t+1} \int E(h_\tau(x))dx \tag{10}$$

subject to $y(\tau) = h_\tau \left(e^{\widehat{v}_1} \ldots e^{\widehat{v}_k} e^{\widehat{\Omega}(\tau-k)} \mu(\tau) \right)$ are considered.

To quantify the discrepancy between the model and the measurements, indicated by $E(h_t)$ above, one can use a discrepancy function f_{in} *inside* the region $g\mu \subset \mathbb{R}^2$, and a discrepancy function f_{out} *outside* the region. These discrepancy functions can be as simple as the indicator functions $f_{in}(x, y) = \chi_y(x) - 1$, and $f_{out}(x, y) = \chi_y(x)$ for the case of binary images representing evolving shapes, or more complex signed-distance scores that can be generalized to grayscale images as shown in [197]. In either case, one can write

$$E(h_\tau) = \int_{g(\tau)\mu(\tau)} f_{in}(x, y(\tau))dx + \int_{g(\tau)\mu^c(\tau)} f_{out}(x, y(\tau))dx. \tag{11}$$

The computation of the gradients $\nabla_\mu \phi$, $\nabla_\Omega \phi$ is described in the next section. Once these gradients have been computed numerically, the general form of the update becomes

$$\begin{cases} \hat{\mu}(t+1) = \hat{\mu}(t) \oplus L_\mu \nabla_\mu \phi \left(\hat{v}(t), \ldots, \hat{v}(t-k), \hat{\mu}(t) \right) \\ \hat{g}(t+1) = e^{\widehat{v}(t)} \hat{g}(t) \\ \hat{v}(t+1) = \hat{v}(t) + L_v \nabla_v \phi \left(\hat{v}(t), \ldots, \hat{v}(t-k), \hat{\mu}(t) \right) \end{cases} \tag{12}$$

where L_μ and L_v are tuning parameters whose effects are discussed in the next section. In the initialization phase, rather than running one step of the gradient above, several steps are run until convergence to a steady state of ϕ. Furthermore, depending on the convergence rate of ϕ relative to the dynamics of μ, it is useful to run several steps of the gradient (even run it to steady state) in the update equation above.

3.5 Experiments

A numerical computation of the gradient above has been used to generate an update for the evolving shape represented implicitly within the level set framework [134]. In particular,

$$\nabla_{\mu}\phi = \sum_{\tau=1}^{k} |J_g(\tau)|(f_{in}(g(\tau)x, y(\tau)) - f_{out}(g(\tau)x, y(\tau)))\vec{N} \tag{13}$$

, where $\vec{N}$ is the outward unit normal and $|J_g(\tau)|$ is the determinant of the Jacobian Matrix of g. The update equation for g, or equivalently Ω, is just the integrator described in the previous section. To update v, one updates each component v_i independently using

$$\nabla_{v_i}\phi = \sum_{\tau=1}^{n} \int_{g^{-1}(\tau)y(\tau)} \langle \frac{\partial g(\tau)x}{\partial v_i}, f_{\{in,out\}}(g(\tau)x, y(\tau))J_{g_*T}(\tau)\rangle ds \tag{14}$$

, where g_* denotes the push-forward and T is the unit tangent vector,

Varying the gain L_μ enforces more or less inertia by μ to change shape. In Figure 7, a five-pixel wide vertical bar has been removed from the images to create an occlusion. The occlusion is close in grayscale value to that of the person being tracked. To see the effect of varying the gain on the estimation of the contour as it passes behind the occlusion, an image sequence with a simple model of motion is used (constant velocity). Therefore a low gain is used on the registration(motion) parameters. As the person passes under the occlusion, the contour will be attracted to the occlusion; without some state estimator, it would grab onto the occlusion. In Figure 7, L_μ is varied. $L_\mu = 0.65$ experimentally turns out to segment the person the best while avoiding getting caught up on the occlusion.

Figure 8 has a total occlusion. Here again the model of motion is certain (constant acceleration) and a low gain on the motion parameters is used. The gains chosen for the highly uncertain contour are $L_\mu = 0.1, 0.5, 0.8$. The initial gain for the contour is set quite low, so measurements are still emphasized, but then the model is able to

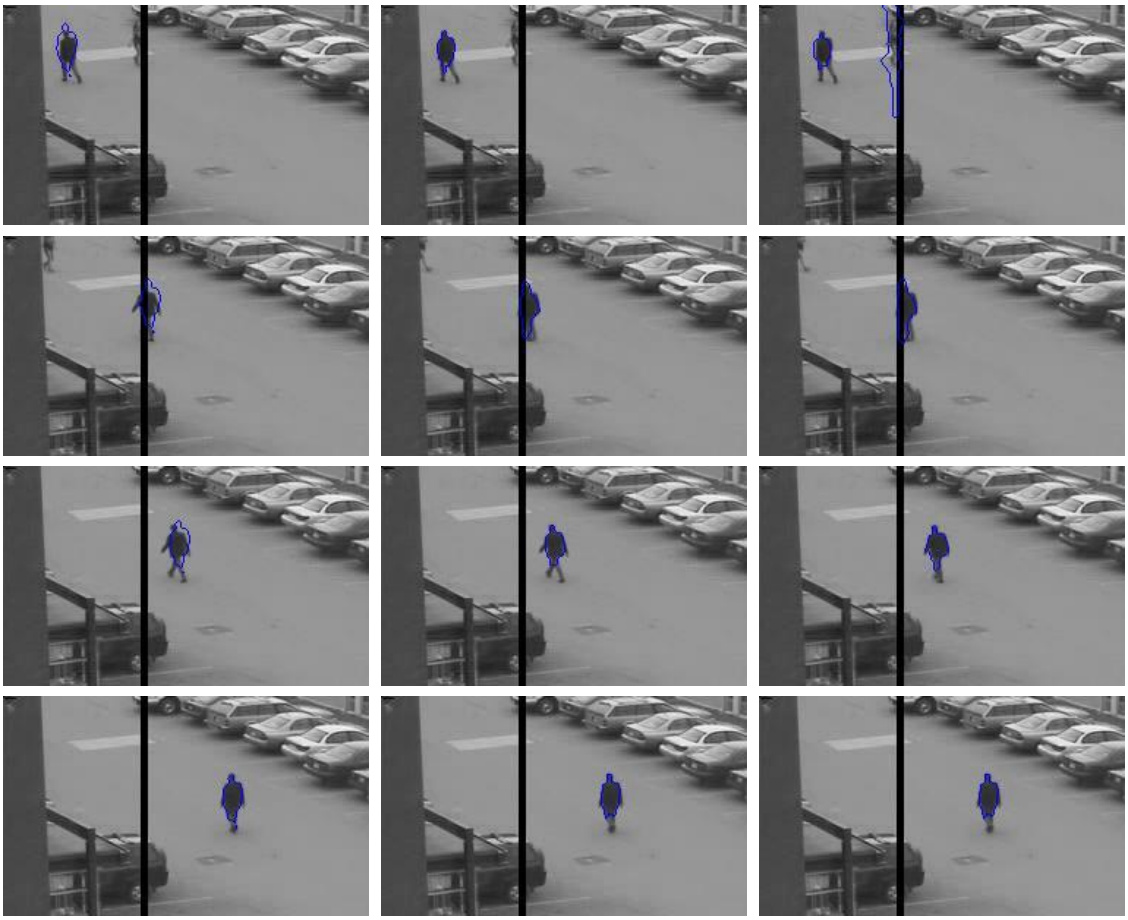

Figure 7: *Tracking a person through an attractive occlusion, Left Column:* $L_\mu = 0.1$, *Middle Column:* $L_\mu = 0.65$, *Right Column:* $L_\mu = 0.7$.

take over and push the tracking of the person through the total occlusion. $L_\mu = 0.1$ tracks the person, but the contour is rather rigid. For $L_\mu = 0.5$, the person is tracked but the contour gets thrown off by the similar-looking books on the printer and poster on the wall. The last example, $L_\mu = 0.8$, emphasizes the measurements too much and loses the person.

In Figure 9, there is only a slight occlusion as the car passes underneath the light pole. But, the occlusion is quite different than the car. The gain L needs to be chosen a bit higher than the previous examples because the car is going into a turn. There is uncertainty in the motion model (constant acceleration) and the segmentation. A low gain (L=0.1 case) on all of the motion parameters and the contour would emphasize the model and keep the contour tracking in the original direction the car was moving. $L = 0.1, 0.3, 0.7$ are examined in this example. The contour is only slightly affected as it passes underneath the pole in the case of $L = 0.3, 0.7$.

3.6 Conclusions on Tracking

This chapter takes a first step in designing a filter for a dynamical model of an evolving contour. The contour is represented implicitly as the (infinite-dimensional) locus of zeros of a given function. That contour evolves in time under the action of a group. Both the underlying state and the group action are estimated from noisy images that can have significant portions of missing data. Although it is hard to say anything analytically about the behavior of such a filter, experiments with challenging real sequences have obtained encouraging results.

Figure 8: *Tracking a person through a severe occlusion, Left Column:* $L_\mu = 0.1$, *Middle Column:* $L_\mu = 0.5$, *Right Column:* $L_\mu = 0.8$.

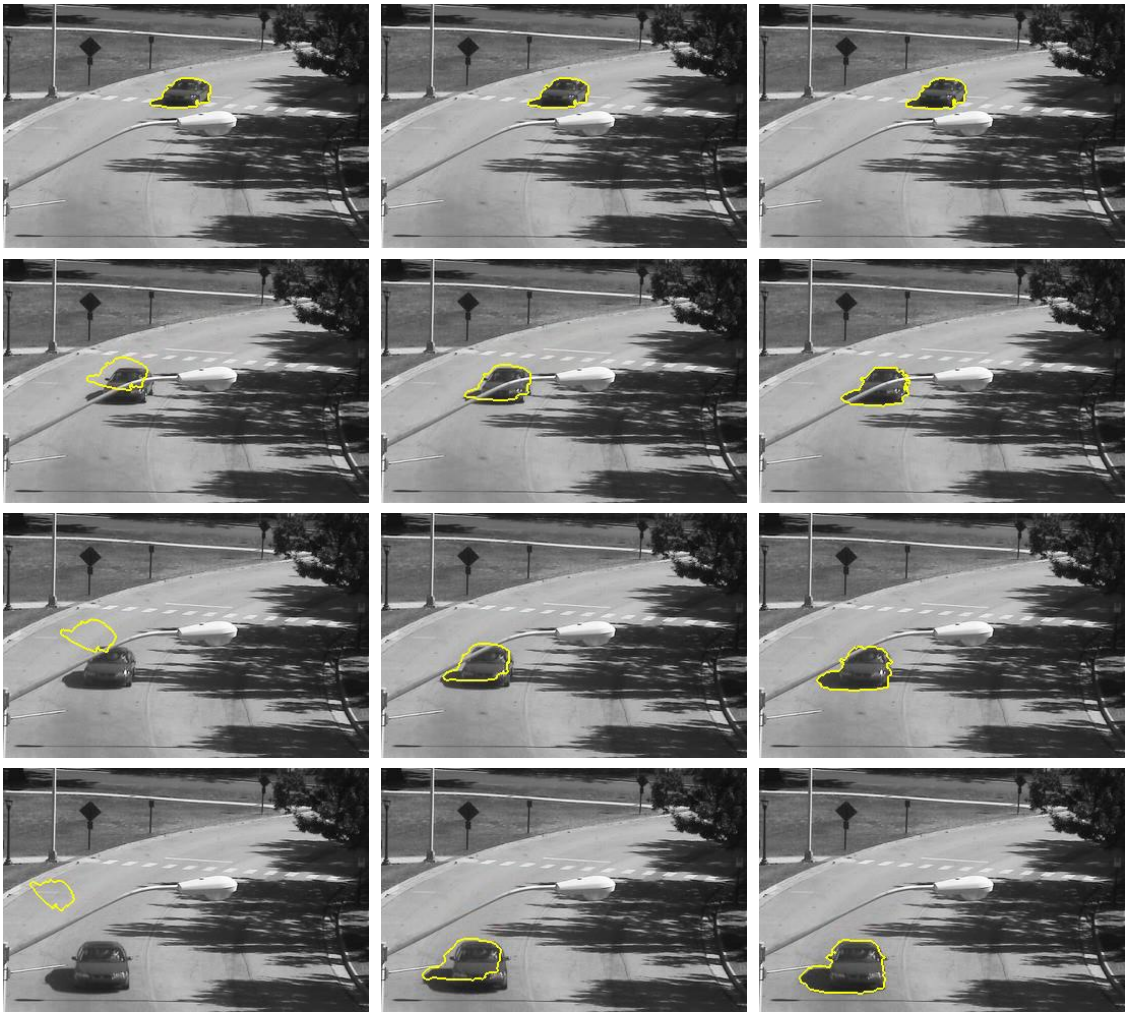

Figure 9: *Tracking a car under a non-attractive occlusion, Left Column:* $L = 0.1$, *Middle Column:* $L = 0.3$, *Right Column:* $L = 0.8$.

CHAPTER IV

LAYERED DEFORMOTION WITH RADIANCE

4.1 Overview

This chapter proposes a model of the shape, motion, and appearance of a scene, seen through a sequence of images which captures occlusions, scene deformations, arbitrary viewpoint variations, and changes in radiance. This model is based on a collection of overlapping layers that can move and deform. Each layer supports an intensity function that can change over time. A discussion follows the generality and limitations of this model in relation to existing models, such as, traditional optical flow or motion segmentation, layers, deformable templates, and deformotion. An illustration is made on how this model is used for inference of shape, motion, deformation and appearance of the scene from a collection of images. The layering structure allows for automatic inpainting of partially occluded regions. The proposed model succeeds on synthetic and real sequences where existing schemes fail, and shows how suitable choices of constants in the model yield existing schemes, from optical flow to motion segmentation and inpainting.

4.2 Background on Layered Deformotion with Radiance

In video sequences changes occur over time because of viewer motion, object motion, or deformation of objects in the scene—including occlusions—and appearance variations arising from the motion of objects relative to the light sources. A suitable model will trade off generality (by allowing variations of shape, motion and appearance) with tractability (by being amenable to inference and analysis). The goal of modeling is to support inference; depending on the application, one may be more

interested in recovering shape (e.g. in shape analysis, classification, recognition, or registration), motion (e.g. tracking, optical flow), or appearance variations (e.g. segmentation), including restoration (inpainting). Traditionally, the modeling task has been approached by making strict assumptions on some of the unknowns to recover the others. Such assumptions include brightness-constancy in optical flow and the affine warping in shape analysis and registration. This approach is partly justified, because in any image-formation model, there is ambiguity between the three factors of shape, motion and appearance; therefore, the most general inference problem is ill-posed. In some applications, for instance video compression, the ambiguity is moot since all that matters is for the model to capture the sequence as faithfully and parsimoniously as possible. Nevertheless, since shape, motion, and appearance affect the generation of the image, a more germane approach would call for modeling all three jointly; then, let complexity dictate the responsibility of each factor, and let the application dictate the choice of suitable regularizers to make the inference algorithms well posed. This research concentrates on modeling, not on any particular application.

A model of image formation is proposed that is general enough to capture shape, motion and appearance variations (Section 4.3) and simple enough to allow inference (Section 4.5). It is desirable to capture *occlusion phenomena*, to entail a notion of *hierarchy* or *layering*, and to capture image variability arising from arbitrary *changes in viewpoint* for non-planar objects. Hence, the model will entail *infinite-dimensional deformations* of the image domain. Such deformations can be due to changes in viewpoint for a rigid scene, or changes of shape of the scene seen from a static viewpoint, or any combination thereof. The model will not attempt to resolve this ambiguity since that requires higher-level knowledge. Furthermore, capturing large-scale motion of objects in the scene, as opposed to deformations, will allow for

a choice of a finite-dimensional group, e.g. Euclidean or affine, separate from infinite-dimensional deformations. An added benefit of this approach is that higher-level knowledge of viewpoint changes may be incorporated through an added prior on this finite-dimensional group to resolve the ambiguity addressed above. Finally, to capture changes in appearance, scene radiance will be one of the unknowns in the model. Changes in radiance can come from changes in reflectance or changes in illumination, including changes in the mutual position between the light sources and the scene. No attempt is made to resolve this ambiguity, since that requires higher-level knowledge.

The image-formation model proposed here is not the most general that one can conceive. Indeed, it is far less general than the simplest models considered acceptable in Computer Graphics, and a lack of generality is illustrated in Section 4.4. Nevertheless, the proposed model is more general than any other model used so far for motion analysis in Computer Vision, as is discussed in Section 4.4, and is complex enough to be barely tractable with the analytical and computational tools available today. The inference problem is posed within a variational formulation, involving partial differential equations, integrated numerically in the level-set framework [134]; although, any other computational scheme of choice would suffice, including stochastic gradients or Markov-chain Monte Carlo. The emphasis of this chapter is to propose a model for shape and appearance of layers and therefore a scene and to show that it can be inferred with at least one particular computational scheme, not to advocate a particular optimization technique.

4.2.1 Relation to Existing Work

This work relates to a wide body of literature in scene modeling, motion estimation, shape analysis, segmentation, and registration. Section 4.4 illustrates the specific relationship between the model proposed and existing models. These existing models include Layers [187, 79], which only model affine deformations of the domain and can

therefore only capture planar scenes under small viewer motion or small aperture, and where there is no explicit spatial consistency within each layer and the appearance of each layer is fixed. The proposed model allows deformations that can model arbitrary viewpoint variation, model layer deformation, and enforce spatial coherence within each layer. One could think of this work as a generalization of existing work on Layers to arbitrary viewpoint changes, or arbitrary scene shape, and changes in radiance (texture); all cast within a variational framework.

The work in this chapter relates to a plethora of variational algorithms for optical flow computation (see [158, 5, 57] and references therein), except that the domain is partitioned and allows arbitrary smooth deformations as well as changes in appearance (which would violate the brightness constancy constraints that most work on optical flow is based on, with a few exceptions, e.g. [77]). It relates to various approaches to motion segmentation, where the domain is also partitioned and allowed to move with a simple motion, e.g. Euclidean or affine, see [51] and references therein. Such approaches do not allow deformations of the region boundaries, or changes in the intensity within each region. Furthermore, they realize a partition, rather than a hierarchy of domain deformations, so this model can be thought of as motion segmentation with moving and deforming layers, changes in intensity, and inpainting [16].

The work in this chapter is closely related to the work of "deformotion" of Yezzi and Soatto [197], except that in this work the layers are allowed to overlap. Thus, this chapter's work can be thought of as a layered version of deformotion with changes in region intensities. The paper by Paragios *et al.* [143], where one distance function is registered to another by using rigid and non-rigid transformations, is also relevant. The work here is also related to the templates [72, 128], in the sense that each of the layers in the given model will be a deformable template. However the shape and intensity profile of the template is estimated along with the layering structure. A

one-layer version of the work here is similar to the paper by Trouve *et al.* [174], where the author describes energies on the manifold $G \times M$, where $g \in G$ is a group action (possibly a C^∞ diffeomorphism or an affine transformation) and M is a manifold consisting of a collection of landmark points or images). With G being the set of C^∞ diffeomorphisms and with M being the set of images, the geodesic between two points $(g_1, m_1), (g_2, m_2) \in G \times M$ describes metamorphoses from one "group-image" pair to another. The active appearance models of [49, 11] have also influenced the approach in this chapter, but here rather than imposing regularization of shape and appearance by projection onto suitably inferred linear subspaces, generic regularizers are employed. Thus this chapter generalizes active appearance models to smooth shape and intensity deformattions, cast in a variational framework.

In the next section the proposed model will be introduced; and in Section 4.5, the approach is illustrated to infer its (infinite-dimensional) constitutive elements.

4.3 Modeling

A scene is represented as a collection of L overlapping *layers*. Each layer, labeled by an index $k = 1, \dots, L$, is a function that has associated with it a domain, or *shape* $\Omega^k \subset \mathbb{R}^2$, and *radiance* function $\rho^k : \Omega^k \to \mathbb{R}^+$. Layer boundaries model the occlusion process, and each layer k undergoes a *motion*, described by a (finite-dimensional) group action g^k, for instance $g^k \in \mathbb{SE}(2)$ (the group of rigid motion on the plane) or the affine group $\mathbb{A}(2)$, and a *deformation*, or *warping*, described by a diffeomorphism $w^k : \Omega^k \to \mathbb{R}^2$, to generate an image I at a given time t. The warping models changes of viewpoint for non-planar scenes, or actual changes in the shape of objects in the scene. Since each image is obtained from the given scene after a different motion and deformation, one indexes each of **the image's corresponding variables** by t: g_t^k, w_t^k, and I_t. Finally, since layers occlude each other, there is a natural ordering in k which, without loss of generality, one can assume to coincide with the integers; layer

$k = 1$ is occluded by layer $k = 2$, and so on. However, since this occlusion model could change (say layer $k = 2$ goes behind layer $k = 3$ and then later layer $k = 2$ is in front of layer $k = 3$), there is a function $l = \max\{k \mid x \in \Omega^k\}$ that indicates the layer that will contribute to the intensity at a pixel in a given image, which is the front most **layer** that intersects the warped domain. Figure 10 shows an example of the layers mapped into an image. For simplicity, assume that $\Omega^0 = \mathbb{R}^2$ (the back

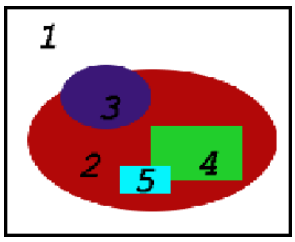

Figure 10: *Labeling of multiple layers of an image.*

most layer, or "the background"). With this notation, the model of how the value of the generic image $I_t : \Omega^0 \to \mathbb{R}^+$ at the location $x \in \Omega^0 \subset \mathbb{R}^2$ is generated can be summarized as $I_t\big(g_t^l \circ w_t^l(x)\big) = \rho^l(x)$, with $x \in \Omega^l$, $l = \max\{k \mid x \in \Omega^k\}$. To simplify the notation, define $x_t^l \doteq g_t^l \circ w_t^l(x)$ (sometimes indicated as x_t for simplicity). The equations describing the generic image are

$$\begin{cases} I_t\big(x_t^l\big) = \rho^l(x), & x \in \Omega^l \\ x_t^l = g_t^l \circ w_t^l(x), & l = \max\{k \mid x \in \Omega^k\}. \end{cases} \tag{15}$$

The goal in this work is to infer, to the extent possible, the radiance family $\{\rho^k\}_{k=1,\dots,L}$, the shape family $\{\Omega^k\}_{k=1,\dots,L}$, the motions $\{g_t^k\}_{k=1,\dots,L;t=1,\dots,N}$, and the deformations $\{w_t^k\}_{k=1,\dots,L;t=1,\dots,N}$ that minimize the discrepancy of the measured images from the ideal model (15), subject to generic regularity constraints. Such a

discrepancy is measured by a cost functional $\phi(\Omega^k, \rho^k, w_t^k, g_t^k)$

$$\phi \doteq \sum_{t=1}^{N} \int_{\Omega^0} \left(I_t(x_t) - \rho^l(w_t^{l-1} \circ g_t^{l-1}(x_t)) \right)^2 dx_t + \zeta \sum_{k=1}^{L} \int_{\partial\Omega^k} ds$$
$$+\lambda \sum_{k=1}^{L} \int_{\Omega^k} \|\nabla \rho^k(x)\|^2 dx + \mu \sum_{k,t=1}^{L,N} \int_{\Omega^l} r(w_t^k(x)) dx \qquad (16)$$
$$\text{subject to} \quad l = \max\{k \mid x \in \Omega^k\}.$$

Here r is a regularizing functional, for instance $r(w) \doteq |\dot{w}| + \frac{1}{|\dot{w}|}$ where $|\dot{w}|$ is the determinant of the Jacobian Matrix of w. Since it is desirable to restrict w to be a one-to-one function, this regularizer r keeps $|\dot{w}|$ close to one. If $|\dot{w}|$ deviates from 1 then either one of the terms $|\dot{w}|$ or $\frac{1}{|\dot{w}|}$ becomes larger. λ, μ, and ζ are positive real constants. Note that l is a function, specifically $l : \Omega^0 \to \mathbb{Z}^+$. The two-norm is chosen for the data-dependent term and the regularizer for simplicity; but other choices could be made. Figure 11 shows the mapping of each layer and the corresponding smooth function.

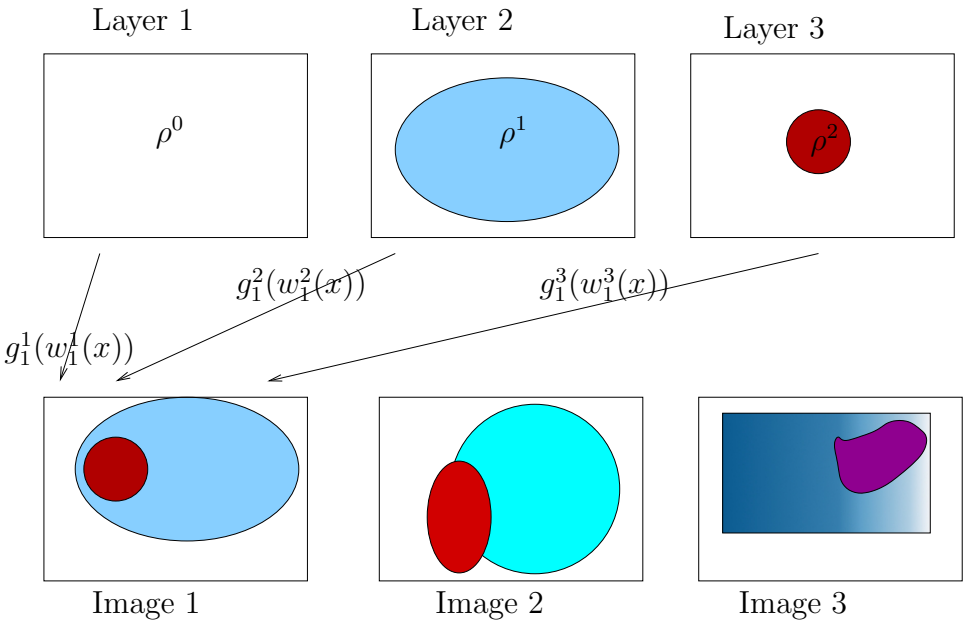

Figure 11: *Multiple layers mapping onto multiple images.*

4.4 Generality of the Model

Equation (15) models images of three-dimensional scenes with piecewise smooth geometry which exhibits Lambertian reflection with piecewise smooth albedo[1] viewed under diffuse illumination from an arbitrarily changing viewpoint. It does not capture global or indirect illumination effects, such as cast shadows or inter-reflections, complex reflectance, such as specularities, anisotropies or sub-surface scattering. These ambiguities are treated as modeling errors and are responsible for the discrepancy between the model and the images, which is measured by ϕ in equation (16). In addition, these discrepancies are lumped together with sensor errors and improperly called "noise." Although far from general, to the knowledge of this writer equation (15) is nevertheless a more ambitious model than has ever been used in the context of motion estimation and tracking. In fact, many existing models are special cases of equation (15).

This model includes traditional **optical flow** as a special case. In particular, if one assumes a single layer to represent the whole image domain (i.e. $L = 0$), a trivial group action (i.e. $g = Id$) and no regularity in the modeled radiance $\rho = \rho^0$ (i.e. $\lambda = 0$) then the resulting minimization problem includes only the radiance ρ and the warps $w_1 = w_1^0$ and $w_2 = w_2^0$ as unknowns (one should consider the case of just two images I_1 and I_2 for now). A much simpler energy is left

$$\phi(\rho, w_1, w_2) = \sum_{t=1}^{2} \int_{\Omega^0} \left(I_t(x_t) - \rho({w_t}^{-1}(x_t)) \right)^2 dx_t + \mu \sum_{t=1}^{2} \int_{\Omega^0} r(w_t(x))\, dx. \qquad (17)$$

If the goal is just to find the warp $w = w_2 \circ w_1^{-1}$ that registers I_1 to I_2 (through the common radiance model ρ), then one may further simplify (17) by setting $w_1 = Id$ and $w_2 = w$, thereby eliminating yet another unknown and yielding (up to a change

[1]The model can be further generalized by allowing ρ^l to be vector-valued to capture a set of radiance statistics such as the coefficients of a filter bank or other texture descriptors, but this is beyond the scope of this chapter.

of measure corresponding to the Jacobian Matrix of w, which is $\dot{w}$)

$$\phi(\rho, w) = \int_{\Omega^0} (I_1(x) - \rho(x))^2 + (I_2(w(x)) - \rho(x))^2 \, dx + \mu \int_{\Omega^0} r(w(x)) \, dx. \tag{18}$$

Since the smoothness penalty on ρ was omitted, it is straightforward to show for a given choice of w that (18) is minimized by setting $\rho(x)$ to the mean of $I_1(x)$ and $I_2(w(x))$. Thus, in this special case (no smoothness on ρ) the joint optimization in (18) may be replaced with a direct optimization of w through this substitution of ρ. The resulting energy

$$\phi(w) = \frac{1}{2} \int_{\Omega^0} (I_1(x) - I_2(w(x)))^2 \, dx + \mu \int_{\Omega^0} r(w(x)) \, dx, \tag{19}$$

depending upon the choice of the regularizer r (note that r typically depends on the derivatives of w rather than its direct values), corresponds to either the classical optical flow in [78] or to one of its many variants.

This model has the advantage of not enforcing global regularization (regularization is imposed within layers, but not across layers), of not comparing images to each other, but to an underlying model (this carries significant advantages when it comes to robustness to noise, as is illustrated in the experiments), and of having an explicit model of the appearance of the scene, which allows "inpainting" individual layers while preserving their motion boundaries.

Choosing $L = 1,\ w = Id,\ \lambda = 0,\ \mu = 0$ yields **motion segmentation,** that has also been addressed by many; see for instance [51] and references therein for the case of affine motion $g \in \mathbb{A}(2)$. In motion segmentation, one partitions the domain into a number of individually moving segments, each of which is assumed to move with a constant (finite-dimensional) motion. Like in optical flow, there is no model of appearance, and the data-dependent term consists of the brightness constancy constraint which forces direct image-to-image comparison:

$$\phi(g_t, \Omega_t) = \sum_t \left(\int_{\Omega_t} (I_t(x) - I_{t+1}(g_t x))^2 dx + \zeta \int_{\partial\Omega_t} ds \right) \tag{20}$$

Note that, in this case, Ω_t is one of the unknowns since w_t is no longer part of the inference, although one could easily define $\Omega_t \doteq w_t(\Omega_0)$, as has been discussed in the previous section.

Choosing $L = 1, \rho = \text{const}, \ r(w) = \|w\|$ yields a model called **deformotion** [197] and has also been extended to grayscale images. The work here is the natural extension of deformotion to layers.

Choosing $L > 1, \ w = Id, \Omega^k$ unconstrained and $g \in \mathbb{A}(2)$ would yield a variational version of the **Layers** model [187], that has never been attempted. It should be noted that this is different than simpler variational multiphase motion segmentation. In the former case the motion of a phase affects the shape of neighboring phases, whereas in the model (15) layers can overlap without distorting underlying domains. One can think of the Layer model as a multiphase motion segmentation with inpainting [16] of occluded layers and shape constraints.

The model also relates to **deformable templates**, where $\rho = \text{const}$ in the traditional model [72] and $\rho = \text{smooth}$ in the more general version [128]. Another relevant approach is **Active Appearance Models** where the regions, warping and radiances are modeled as points in a linear space.

$$w_t^k(x) = w_0^k(x) + W^k(x)s_t \tag{21}$$

where $w_0 : \Omega^k \to \mathbb{R}^2$ and $W^k : \Omega^k \to \mathbb{R}^n$ denotes a set of basis functions or principal components, and $s_t \in \mathbb{R}^n, \ t = 1, \dots, N$ is a vector of shape coefficients. Similarly,

$$\rho^k(x) = \bar{\rho}^k(x) + P^k(x)\alpha^k \tag{22}$$

where $\bar{\rho}^k : \Omega^k \to \mathbb{R}$ and $P^k(x) : \Omega^k \to \mathbb{R}^n$ is a vector of principal components, and $\alpha^k \in \mathbb{R}^n$ a vector of appearance parameters. It should be noted that the functions P^k and W^k have to satisfy orthogonality constraints, and these have to be enforced during the inference of the bases. The model (15) does not impose such restrictions and renders the problem well-posed by generic regularization instead.

Finally, by virtue of the regularization imposed on ρ, the scheme in this chapter relates to **image inpainting**, except that one performs inpainting *both* by layer transfer from multiple images and by regularization. The advantage of the method in this chapter is that it can exploit whatever information is there: If multiple views are available, their contribution is weighted relative to the harmonic interpolation term. If only one image is available, then intensity regularization dictates the filling process.

4.5 Inference

Minimizing the cost functional in (16) is a tall order. The functional depends upon each domain Ω^k and its boundary (a closed planar contour), its deformation (a flow of planar diffeomorphisms) w_t^k, and its radiance (a piecewise smooth function) ρ^k—all of which are infinite-dimensional unknowns. In addition, the functional depends on a group action per layer per instant, g_t^k, on the occlusion model, which is represented by the discrete-valued function $l(x) = \max\{k \mid x \in \Omega^k\}$, for each layer $k = 1, \ldots, L$.

As long as each layer is a compact region bounded by a simple smooth curve, there is no loss of generality in assuming that the Ω^k are fixed, because each diffeomorphism w_t^k will act transitively on Ω^k. Therefore, it can be assumed that each region Ω^k is a circle in some of the examples. While there is no loss of generality, there is a loss of energy; if one were allowed to also optimize with respect to the initial regions, one would be able to reach each deforming layer faster and with less energy. This, however, does not enhance the generality of the model, hence it will be forgone for some examples.

Apart from this simplification, one proceeds by minimizing the functional in equation (16) by using simultaneous gradient flows with respect to the groups (motion), the radiances (appearance), and the diffeomorphisms (deformation). The detailed evolution equations are complicated, depending upon the number of layers and the

occlusion structure between layers. To help avoid excessive subscripting and superscripting and multiple-case definitions according to occlusion relationships, the key properties of the various gradient terms will be outlined for the case of a background layer Ω^0, a single image I, and a single foreground layer Ω^1. To keep the illustration simple, it will be assumed that the group action g^0 and the warp w^0 for the background layer are simply the identity transforms. This is the simplest possible scenario that will allow the key properties of the gradient flows to be shown.

Let $\hat{x} = g^1(w^1(x))$ and $\hat{\Omega}^1 = g^1(w^1(\Omega^1))$. With this notation, one may write the image-dependent terms in the energy functional as

$$E = \int_{\hat{\Omega}^1} \big(I(\hat{x}) - \rho^1(x)\big)^2 \, d\hat{x} + \int_{\Omega^0 \setminus \hat{\Omega}^1} \big(I(x) - \rho^0(x)\big)^2 \, dx. \tag{23}$$

If g denotes any single parameter (e.g. horizontal translation) of the group g^1, then differentiating yields

$$\begin{aligned} \frac{\partial E}{\partial g} = &\int_{\partial\hat{\Omega}^1} \left\langle \frac{\partial \hat{x}}{\partial g}, \hat{N} \right\rangle \Big(\big(I(\hat{x}) - \rho^1(x)\big)^2 - \big(I(\hat{x}) - \rho^0(\hat{x})\big)^2 \Big) \, d\hat{s} + \\ &+ 2 \int_{\hat{\Omega}^1} \big(I(\hat{x}) - \rho^1(x)\big) \left\langle \nabla \rho(x), \operatorname{inv}\left[\big(w^1\big)' \right] \frac{\partial}{\partial g} \operatorname{inv}[g^1](\hat{x}) \right\rangle d\hat{x}, \end{aligned} \tag{24}$$

where $\hat{N}$ and $d\hat{s}$ denote the outward unit normal and the arclength element of $\partial\hat{\Omega}^1$, respectively. For a multidimensional group, the procedure can be repeated for each parameter in the local coordinate representation of the group.

The update equations for the group involve measurements both along the boundary of its corresponding layer (first integral) as well as measurements within the layer's interior (second integral). This latter integral vanishes if a constant radiance ρ is used for the layer. It is not necessary to differentiate the image data I; derivatives land on the estimated smooth radiance ρ instead, which is a significant computational perk of the model in this chapter that results in considerable robustness to image noise [196].

A similar gradient structure arises for the case of the infinite dimensional warp w; the boundary-based terms and region-based terms for each layer are similar to

previous integrals. However, additional terms arise in the gradient flow equations for w depending upon the choice of regularization terms in the energy functional, such as smoothness penalties, and magnitude penalties.

To solve for the transformations of layer 1, the superscript is dropped on w and g. Let $w(x) = [x + u(x), y + v(x)]^T$. To solve for u and v at time n, one may use the following explicit iterative method:

$$\begin{bmatrix} u^n(x) \\ v^n(x) \end{bmatrix} = \begin{bmatrix} u^{n-1}(x) - dt * (\delta(\partial\Omega^1) * u_c^{n-1}(x) + u_r^{n-1}(x) - \Delta u^{n-1}(x)) \\ v^{n-1}(x) - dt * (\delta(\partial\Omega^1) * v_c^{n-1}(x) + v_r^{n-1}(x) - \Delta v^{n-1}(x)) \end{bmatrix} \tag{25}$$

where

$$\begin{bmatrix} u_c(x) \\ v_c(x) \end{bmatrix} = S^T R^T J g' w'(x) J^T \begin{bmatrix} N^x(x) \\ N^y(x) \end{bmatrix} \left[(I \circ g \circ w(x) - \rho^1(x))^2 - (I \circ g \circ w(x) - \rho^0(x))^2 \right] \tag{26}$$

and

$$\begin{bmatrix} u_r(x) \\ v_r(x) \end{bmatrix} = -2|g'||w'(x)| \begin{bmatrix} (I \circ g \circ w(x) - \rho^1(x))(-\rho_x^1(x)) \\ (I \circ g \circ w(x) - \rho^1(x))(-\rho_y^1(x)) \end{bmatrix} \tag{27}$$

Here $R = \begin{bmatrix} c\theta & s\theta \\ -s\theta & c\theta \end{bmatrix}$ is the rotation matrix, $S = \begin{bmatrix} xs & 0 \\ 0 & ys \end{bmatrix}$ is the scaling matrix and $J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is the 90° rotation matrix. The curve evolution is also similar to the boundary-based term for the evolution of g:

$$\frac{\partial C}{\partial t} = -\Big(\big(I(\hat{x}) - \rho^1(x)\big)^2 - \big(I(\hat{x}) - \rho^0(\hat{x})\big)^2 \Big) \hat{N}. \tag{28}$$

Finally, the optimality conditions for the smooth radiance functions ρ^0 and ρ^1 are given by the Poisson-type equations

$$\Delta\rho^1(x) = \lambda\big(\rho^1(x) - I(\hat{x})\big), \qquad x \in \Omega^1 \tag{29}$$

$$\Delta\rho^0(x) = \begin{cases} 0, & x \in \hat{\Omega}^1 \\ \lambda\big(\rho^0(x) - I(x)\big), & x \in \Omega^0 \setminus \hat{\Omega}^1. \end{cases} \tag{30}$$

Notice that the background radiance ρ^0 is "inpainted" in regions occluded by the foreground layer Ω^1 by harmonic interpolation from the boundary of $\hat{\Omega}^1$, since ρ satisfies Laplace's equation $\Delta\rho^0 = 0$. Once all the terms are put together, one can generate a gradient flow that simultaneously evolves all layer assignments, boundaries, and intensities. In the next section some of the features of the model and the resulting optimization will be illustrated and compared to existing schemes.

For a single parameter g from the mapping g_t^k from layer k to an image t, the full evolution of the parameters g are governed by

$$\frac{\partial E}{\partial g} = \int_{\partial\Omega_t^k} \delta_l(k, x_t) \left\langle \frac{\partial x_t}{\partial g}, N_t \right\rangle \Big(\big(I(x_t) - \rho^k(x)\big)^2 - \big(I(x_t) - \rho^m(x)\big)^2 \Big) ds_t + \tag{31}$$
$$+2 \int_{\Omega_t^k} \delta_l(k, x_t) \big(I(x_t) - \rho^k(x)\big) \left\langle \nabla\rho^k(x), \text{inv}\left[\left(w_t^k\right)'\right] \frac{\partial}{\partial g} \text{inv}[g_t^k](x_t) \right\rangle dx_t,$$

where $\delta_l(k, x_t)$ is 1 when $l(x_t) = k$ and 0 otherwise, and $m = l - 1$ when $k = l$.

4.6 Experiments

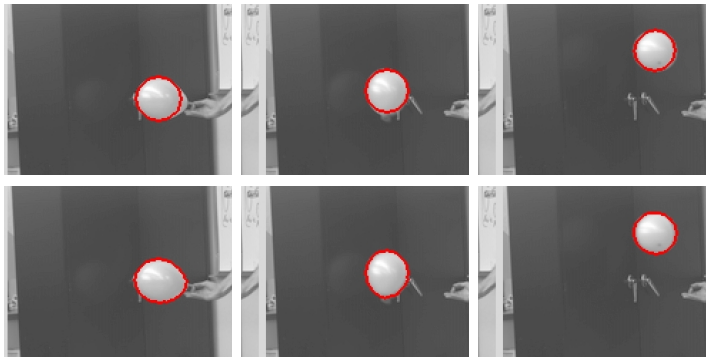

Figure 12: *Tracking a balloon: Three sample views are shown from a sequence of a deflating balloon moving with an erratic motion while changing its shape from a drop-like shape to a circle. In the top row the boundary of the first layer is shown as estimated by a rigid layer model with a single scaling term that does not allow for layer deformation, akin to a variational implementation of traditional layer models. The model tracks the motion of the layer, but it fails to capture its deformation. On the bottom row the same three images are shown with the first layer superimposed, where the layer is allowed to both move (rigidly) and deform (diffeomorphically), yielding 82 percent lower RMS residual error, and capturing the subtler shape variations.*

The capability of the model in this chapter to track deforming layers is illustrated in the first experiment. In Figure 12, three images of a sequence show a deflating balloon undergoing a rather erratic motion while deforming from an initial waterdrop shape to a circular one, and finally to a drop-like shape. On the top row of Figure 12, the layer boundaries are shown for a model that only allows for rigid deformations of the initial contour (a circle) using a single scaling term—this is essentially a variational implementation of the model of [187]. It captures the gross motion of the balloon, but it cannot capture the subtler shape variations. The second row shows the same three sample images with the boundary of the first layer superimposed, where the layer is allowed to deform according to the model introduced in this chapter. The data fidelity term used is a Mumford-Shah term so the radiances representing each layer are smooth functions. The layer changes shape to adapt to the deforming balloon, all while capturing its rather erratic motion. The average root mean squared (RMS) error per image for the affine layer model is 30.87, whereas the residual for the case of the deforming layers is 5.51. More importantly, the phenomenology of the scene, visible in the figure, has been correctly captured.

In the next experiment, all the features of the model of this chapter are illustrated by showing how the model recovers the background behind partially occluded layers while recovering their motion and deformation. Figure 13 shows a few samples from a dataset where the silhouette of a moving hand forms a victory sign while moving the relative position between the fingers. The background, which is partially occluded, is a spiral. The average shape is used as the initial shape of the foreground layer to find its affine motion and the diffeomorphic warp w_i. The radiance within each layer is assumed to be smooth, so when the background layer is recovered, a slightly smoothed version of the spiral is shown. Of course, one could further segment the black spiral from the background to obtain sharp boundaries, but this would not help to illustrate the feature of the model, therefore it is not illustrated here.

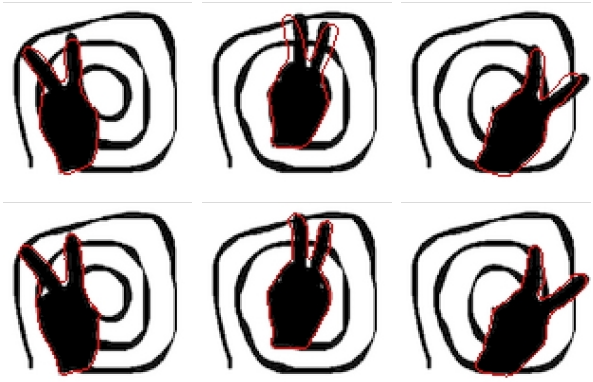

Figure 13: *Victory sign, with deforming hand, moving in front of a partially occluded background portraying a spiral. The goal here is to recover the radiance of each layer (the spiral in the background and the constant black intensity of the hand), as well as the motion and deformation of the foreground layer. Note that layer models based only on affine motion would fail to capture the phenomenology of this scene by over-segmenting the region into three regions, each moving with independent affine motion. The model captures the overall motion of the layer with an affine group, and then the relative motion between the fingers as a deformation, as illustrated in Figure 14.*

Figure 14 summarizes the modeling process. The recovered layers are shown on the top row. Since the radiance is assmued to be smooth within each layer, one can only recover a smoothed version of the spiral. These layers are deformed according to a diffeomorphism; one per layer, each defined on the domain of the layer (second row) and then moved according to an affine motion. The third row shows the image generated by the model which can be thought of as a deterministic generative model since it performs comparisons at the image level, not via some intermediate representation. The corresponding images are displayed in the last row with the layers superimposed for comparison.

The next set of experiments uses standard sequences used for optical flow analysis and is designed to illustrate the difference between the model of this chapter and standard optical flow. A representative set of results of the motion field estimated by optical flow (left) and the model of this chapter (right) is reported in Figure 16. The model in this chapter does not rely on global regularization, but only regularization within each layer segmented in Figure 15. Therefore, the boundaries of the motion

field are better resolved.

Naturally, the model in this chapter is a superset of those commonly used for optical flow computation. This point is illustrated by reducing the weight of the smoothness term for ρ in Figure 17, which yields results closer to standard optical flow. In comparison to the ground truth vector field, the vector field given by optical flow has an average angular error of 8.12°. Deformotion with a smoothness weight λ of 200 gives an average angular error of 9.99°. Reducing the smoothness weight λ to 20 gives an average angular error of 8.11°, which is closer to the result of optical flow.

There can be some benefit in changing $r(w)$ from an L2 type norm to an L1 norm, which has been shown in Papenburg *et al.* [135] to improve optical flow. Instead of an L2 norm

$$r(w) = \int \langle \nabla u, \nabla u \rangle + \langle \nabla v, \nabla v \rangle \ dx, \tag{32}$$

an L1 norm is used

$$r(w) = \int \sqrt{\langle \nabla u, \nabla u \rangle + \langle \nabla v, \nabla v \rangle + \epsilon^2} \ dx \tag{33}$$

where ϵ is small, here chosen to be $\epsilon = .001$. Figure 18 shows a comparison using the two norms combined with deformotion and optical flow.

A beneficial side-effect of having an explicit model of the scene, i.e., a regular irradiance pattern, with smoothness controlled by λ, is the possibility of comparing individual images to a (noiseless) model rather than comparing noisy images to each other. This benefit is visible in Figure 19, where the flow field obtained with the layered deformotion model with L1 on artificially corrupted sequences is closer to the cleaner version of the sequence than using L1 standard optical flow.

The comparison with optical flow illustrates the necessity for partitioning the domain into independently moving objects. This is a motion segmentation task. Therefore, the model of this chapter is compared with more standard ones that partition the flow into affine segments, while still relying on the brightness constancy constraint

and without using an explicit model of the appearance of the scene. Such models can be obtained by increasing the regularization of the layer deformation (i.e. the entire layer moves with the same finite-dimensional motion: translational, Euclidean or affine). Figure 20 illustrates this effect.

Note that the model in this chapter, by virtue of having an explicit representation of the appearance of each layer, can automatically fill in the appearance of underlying layers, as illustrated in Figure 21.

Figure 22 illustrates inpainting using the model of this chapter. In this example, there is some camera motion which makes the whiteboards in the two images not quite lined up. Also, there has been some corruption of the images which is modeled as the foreground layer that is moved around via an affine group. The whiteboard (background layer) is recovered with its own affine registration and the inpainted whiteboard is shown.

The conclusion to draw from these experiments is that the model in this chapter, a superset of existing schemes (optical flow, motion segmentation, deformotion, and inpainting), allows the user to apply existing algorithms by using proper choices of constants. Naturally, the price to pay for such flexibility and for the added power stemming from a richer model is computational complexity. All the experiments shown have been run on a pentium M 2GHz PC and take five minutes per 1000 iterations.

4.7 Discussion

This chapter presents a generative model of the appearance (piecewise smooth albedo), motion (affine transformation) and deformation (diffeomorphism) of a sequence of images that exhibit occlusions. This model is used as a basis for a variational optimization algorithm that simultaneously tracks the motion of a number of overlapping layers, estimates their deformation, and estimates the albedo of each layer, including

portions that were partially occluded. Where no information is available, the layers are implicitly inpainted by their regularizers.

This model generalizes existing layer models to the case of deforming layers. Alternatively, one can think of the algorithm as a layered version of deformable tracking algorithms or as a generalized version of optical flow or motion segmentation where multiple layers are allowed to occlude each other without disturbing the estimate of adjacent and occluded layers.

The numerical implementation of the method in this chapter uses level set methods. It is realized without taking derivatives of the image, a feature that yields significant robustness when compared with boundary-based approaches to estimating optical flow. The approach has been illustrated on simple but representative sequences where existing methods fail to capture the phenomenology of the scene by either over-segmenting it, or by failing to capture its deformation while only matching its affine motion.

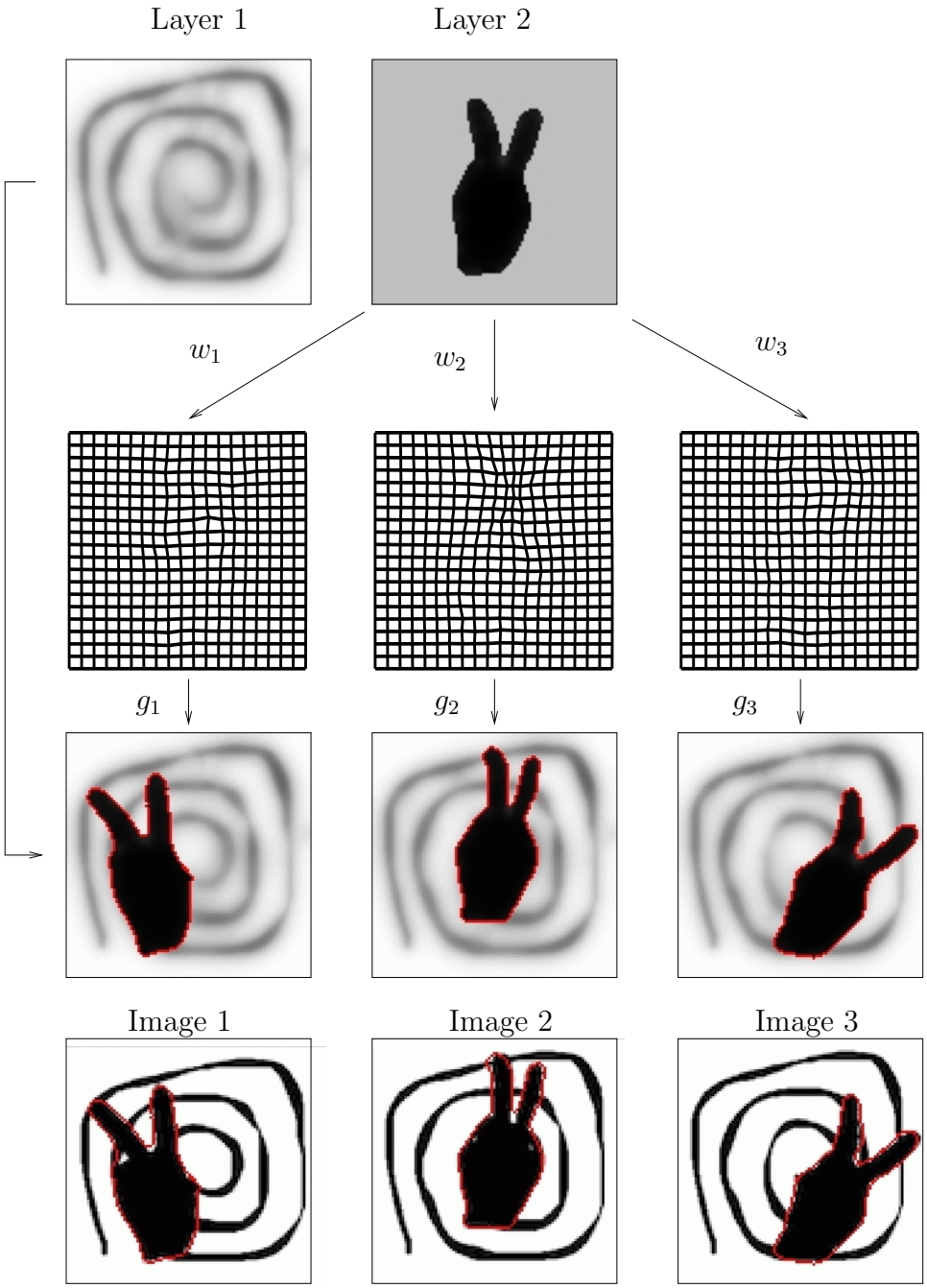

Figure 14: *Multiple layers mapping onto multiple images: The inference process returns an estimate of the albedo in each layer (top). Since smooth albedo is assumed, the spiral is smoothed. The deformation of each layer is estimated (second row), together with its affine motion, to yield an approximation of the image (third row). This image approximation is used for comparison with the measured images (bottom row) that drives the optimization scheme.*

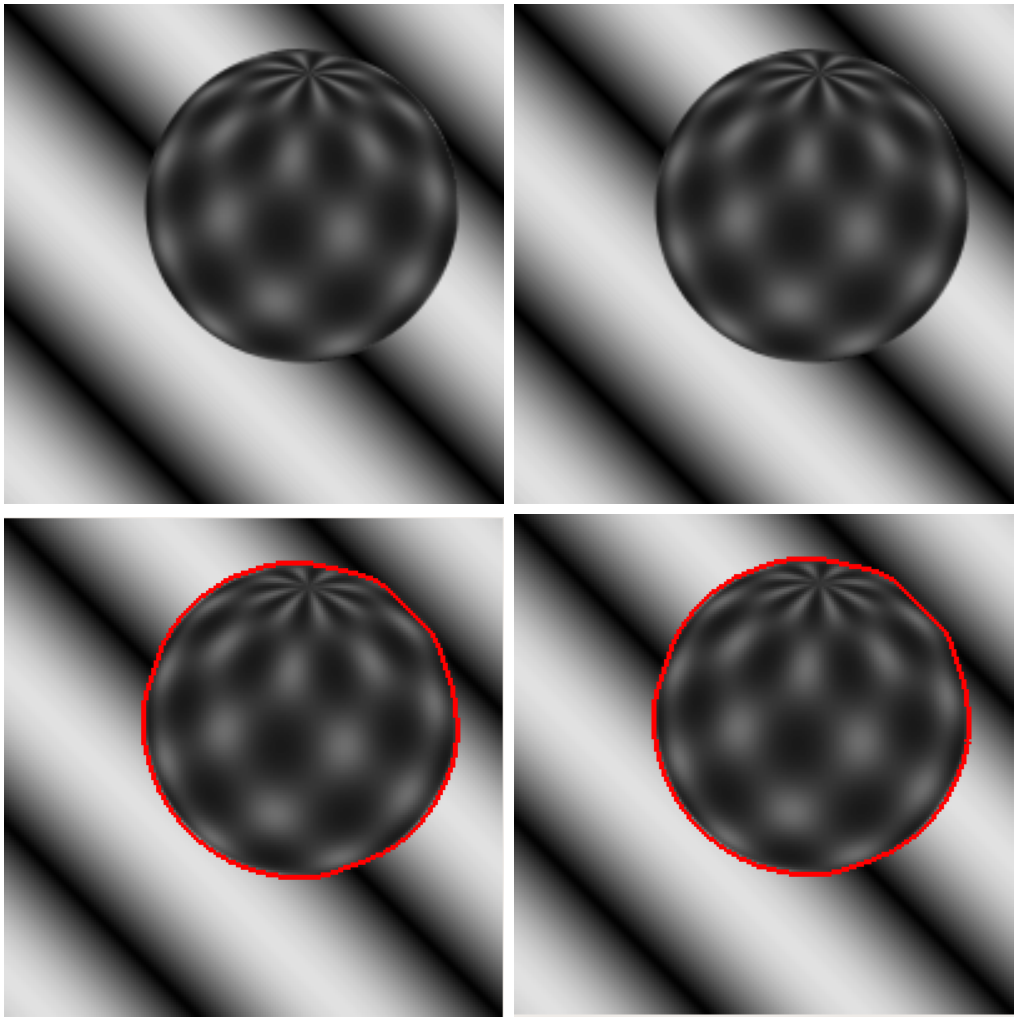

Figure 15: *Rotating sphere and segmentation obtained using deformotion.*

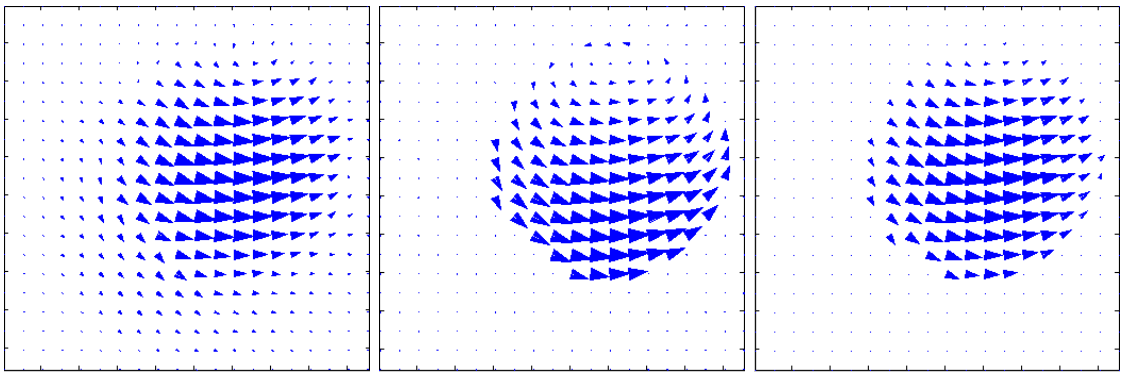

Figure 16: *Optical flow; ground truth; deformotion: Standard optical flow (left) imposes global regularization, which results in errors at the boundary (the vector field is more spread out than the one from the proposed model, on the right). The ground truth is in the middle. The average angular errors for optical flow and deformotion are* 11.49° *and* 6.31° *respectively. The standard deviation for the angular errors are* 1.37° *and* 1.44°*. The parameters and regularization constants used were* dt= 0.2, *iterations*= 10000, $\alpha = 10$ *(data fidelity),* $\mu = 0.5$*, (smoothness of* w*),* $\lambda = 5$ *(smoothness of* ρ*).*

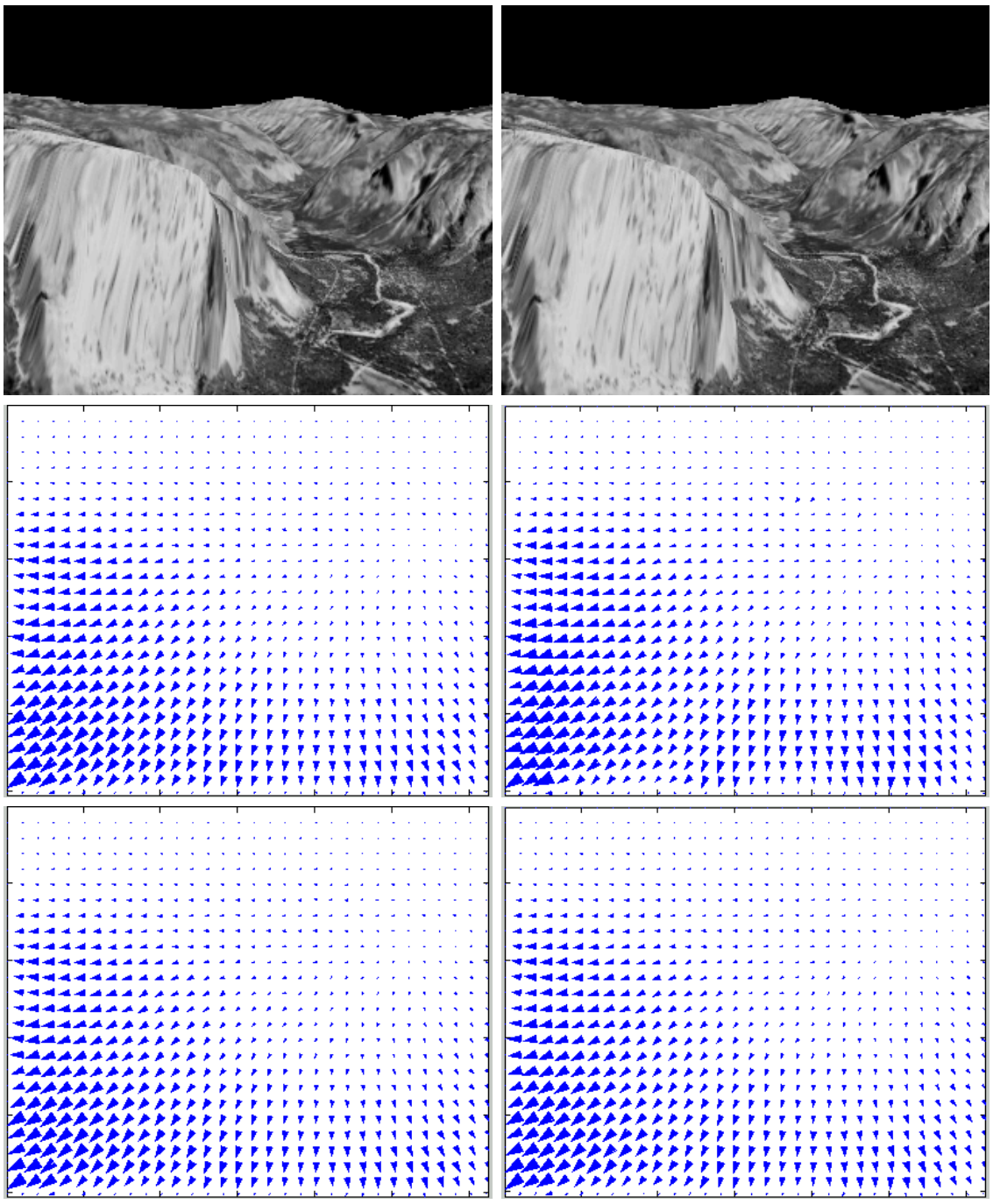

Figure 17: *Deformotion to optical flow: Optical flow (left) can be obtained from the general model (right) by allowing* $\lambda \to 0$. *Compare the results with parameters* dt= 0.028, *iterations*= 71000, $\alpha = 20$, $\mu = 0.55$, *and* $\lambda = 20$ *on the bottom row with* 200 *on the middle row. Note that the two models (left and right) are closer on the bottom row. In comparison to the ground truth vector field, the vector field given by optical flow has an average angular error of 8.12. Deformotion with a smoothness weight of* 200 *gives an average angular error of 9.99. Reducing the smoothness weight to* 20 *gives an average angular error of 8.11, which is closer to the result of optical flow.*

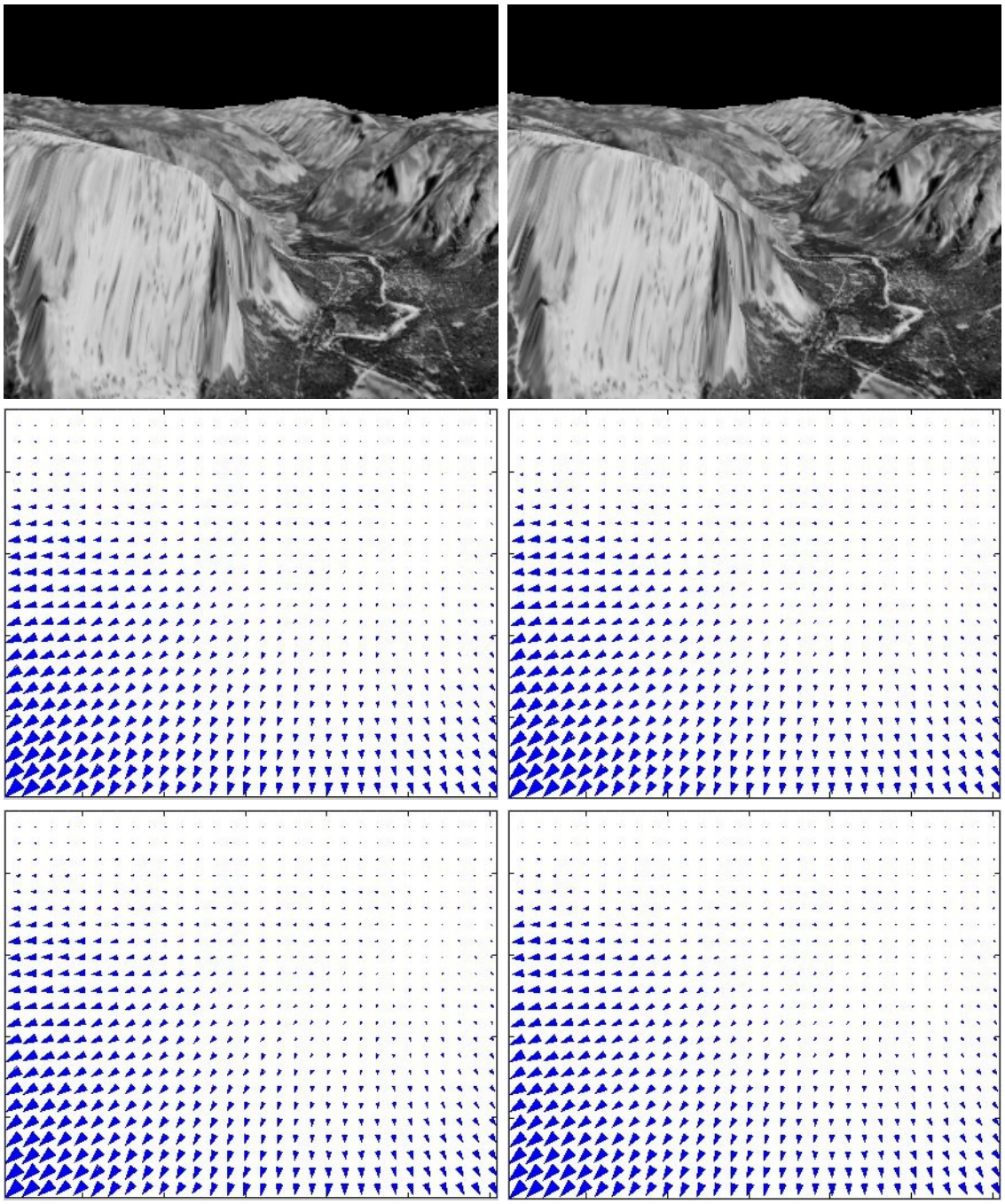

Figure 18: *L1 deformotion; L2 deformotion; optical flow: L2 is on the middle row with deformotion first and optical flow second. The average angular errors are* 4.83° *and* 4.74° *respectively. L1 is on the bottom row with average angular errors of* 4.80° *and* 4.67°. *The parameters used are dt*= 0.1, *iterations*= 20000, $\alpha = 1$, $\mu = 500$, $\lambda = 1$.

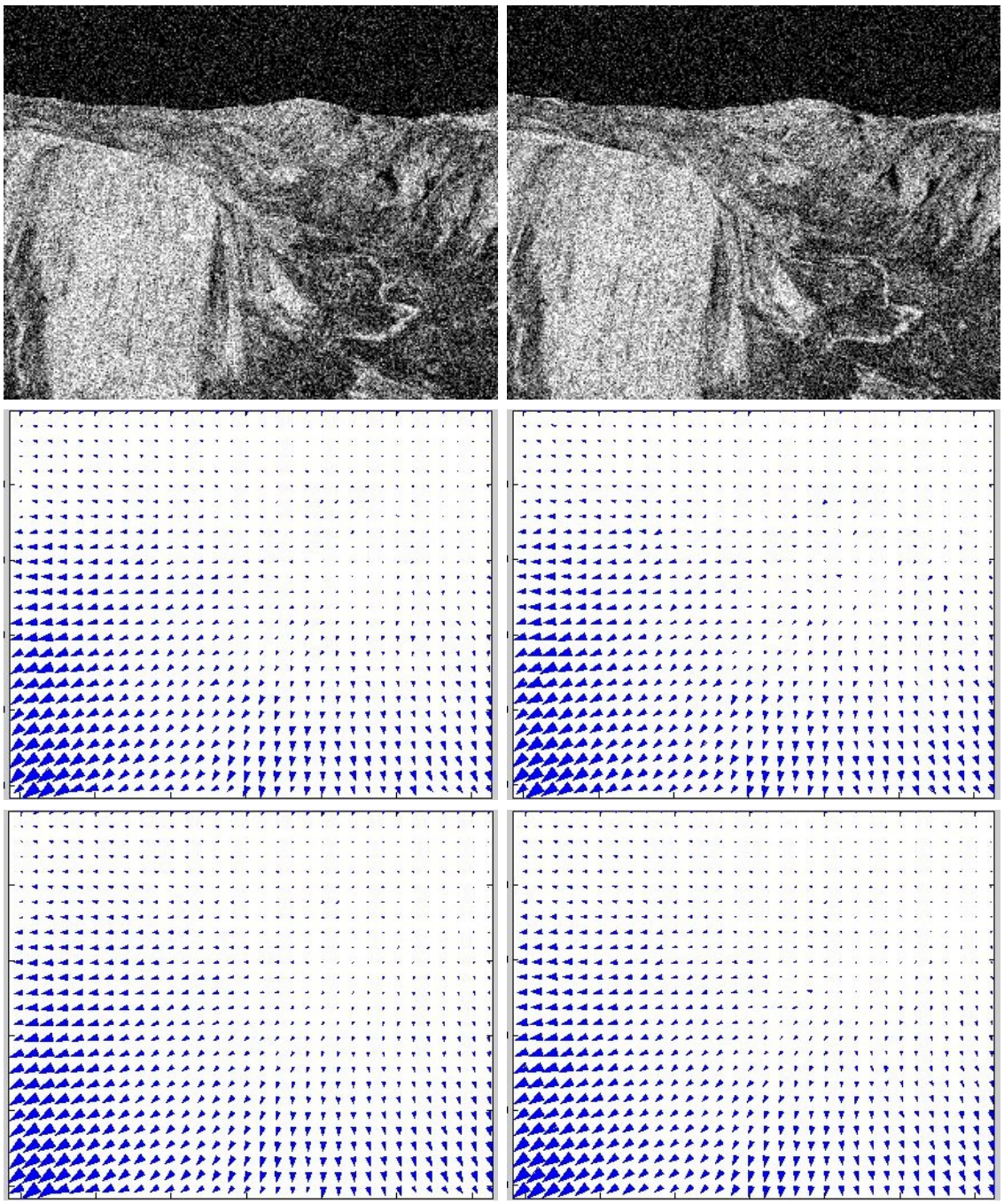

Figure 19: *Noisy Case: L1 deformotion; L2 deformotion; optical flow. The images have been corrupted with Gaussian noise of zero mean and a variance of 0.05. L2 is on the middle row with deformotion first and optical flow second. The average angular errors are* 11.93° *and* 13.23° *respectively. L1 is on the bottom row with average angular errors of* 11.26° *and* 11.88°*. The L1 version of deformotion attains the best result with the angular error of* 11.26°*. The parameters used are* $dt = 0.1$*,* $iterations = 20000$*,* $\alpha = 1$*,* $\mu = 5000$*,* $\lambda = 30$*.*

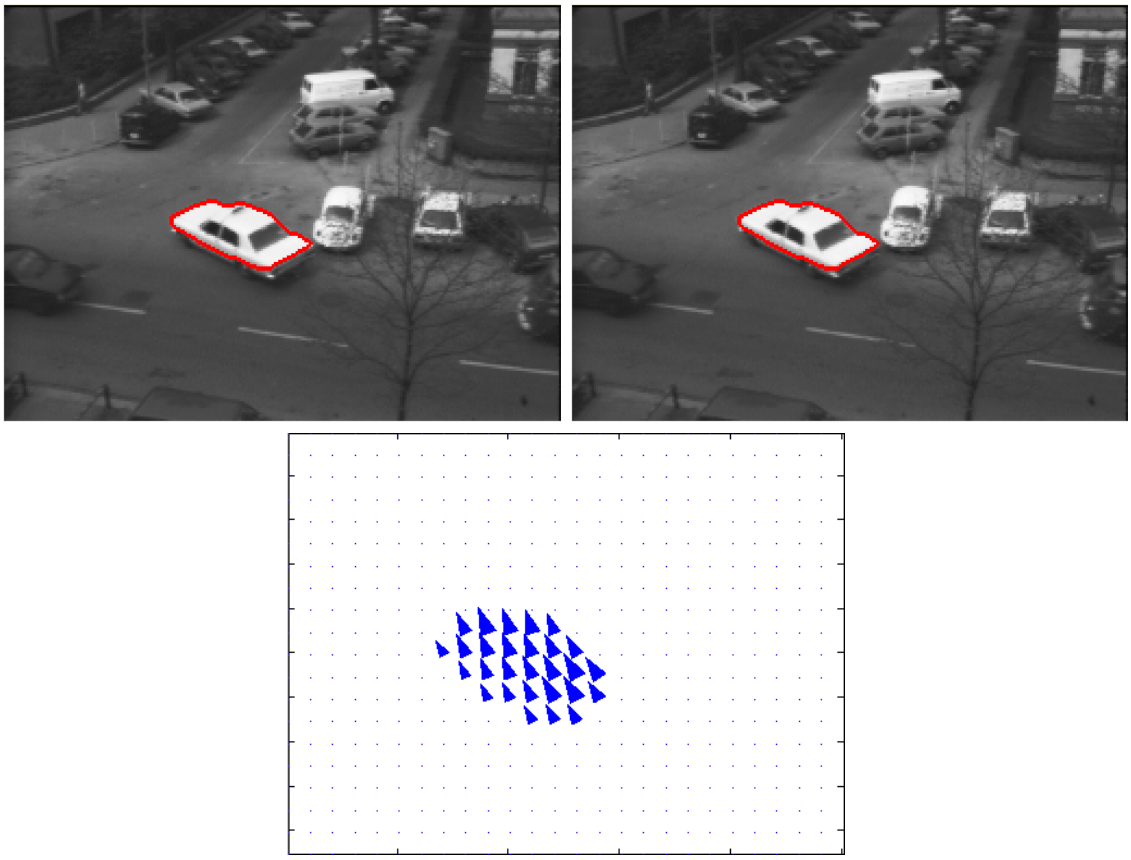

Figure 20: *The model proposed can be used to perform motion segmentation by increasing the regularization μ of the domain deformation for each layer (parameters used: dt= $\frac{0.2}{4000}$, iterations= 20000, $\alpha = 20$, $\mu = 4000$, $\lambda = 2000$).*

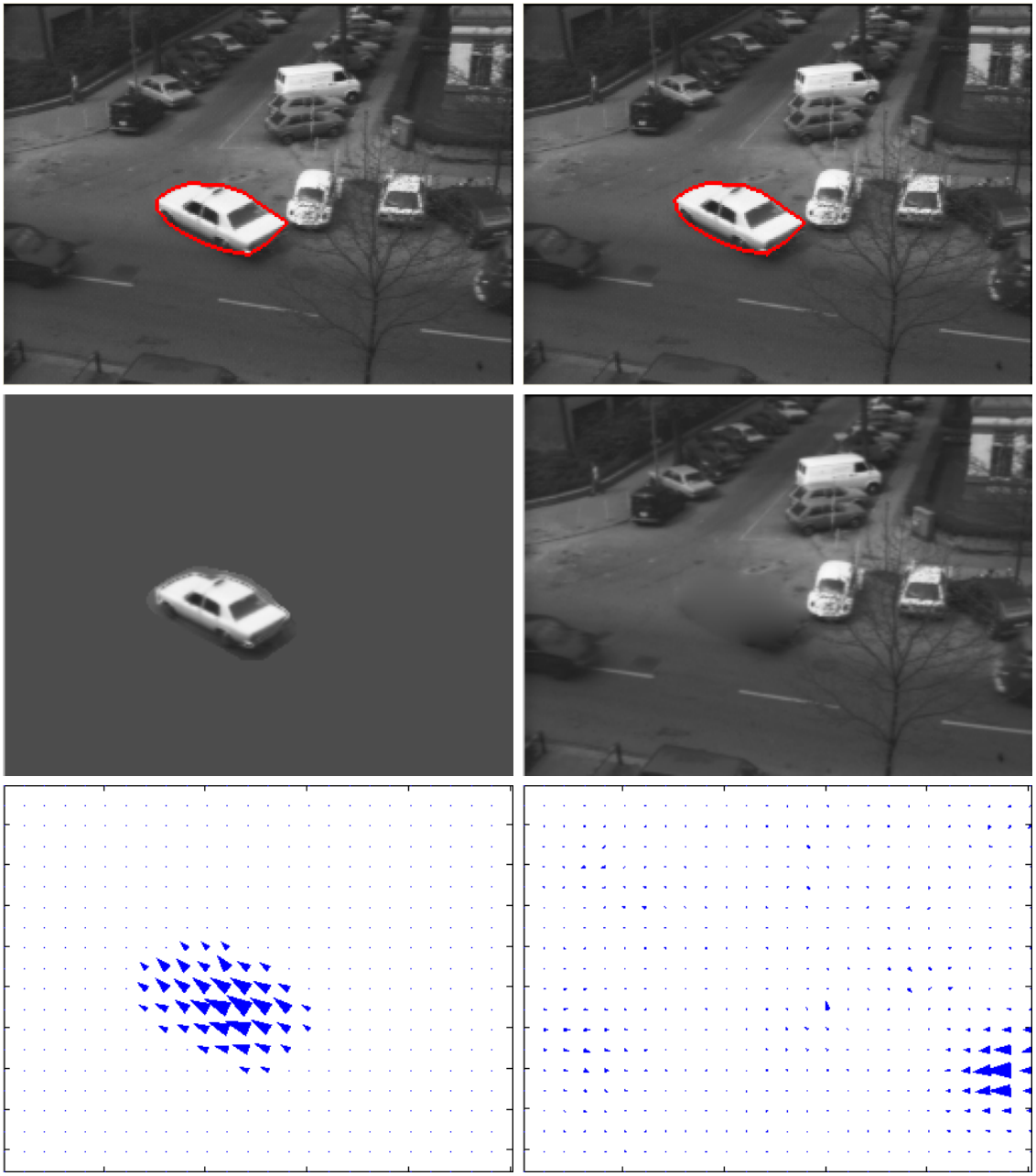

Figure 21: *The model yields "inpainted" layers. The top row shows the boundaries of layers; the middle row, the reconstructed appearance of the layers (ρ); and the bottom row, the warpings (w). Parameters used: dt= 0.2, iterations= 2000, $\alpha = 20$, $\mu = 1.0$, $\lambda = 0.8$, $\zeta = 3.0$ (arclength weight) .*

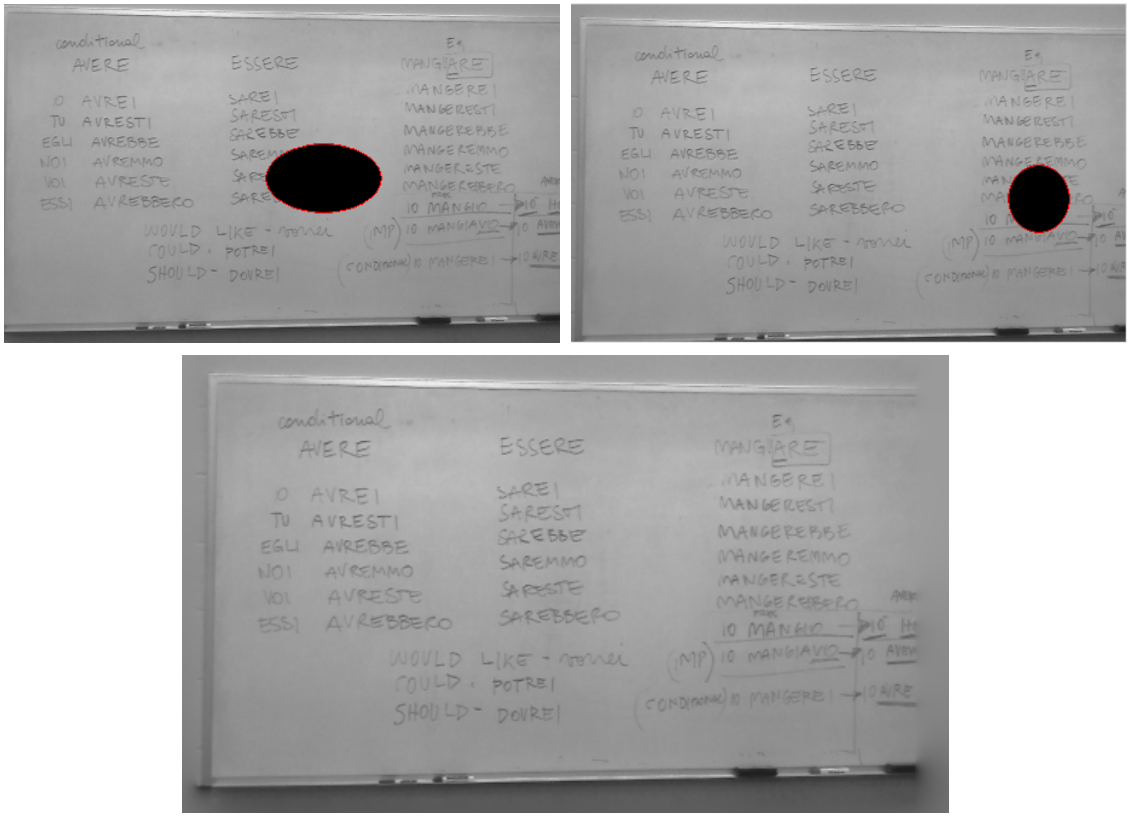

Figure 22: *Image inpainting with the model. First two images: corrupted images of a teacher's whiteboard with some camera jitter. Last image: Image inpainting result.*

CHAPTER V

LAYERED DEFORMOTION WITH A JOINT PRIOR

5.1 Overview

Modeling the variability of different images of the same scene, or class of objects, obtained by changing the imaging conditions; for instance, the viewpoint or the illumination, is challenging. Understanding such variability is key to the reconstruction of objects despite changes in their appearance (e.g. due to non-Lambertian reflection), or to the recognition of classes of objects (e.g. cars), or individual objects seen from different vantage points. This chapter presents a model that can account for changes in shape or viewpoint, appearance, and occlusions of line of sight. The model learns a prior model for each factor (shape, motion and appearance) from a collection of samples using principal components analysis, akin to a generalization of "active appearance models" to dense domains affected by occlusions.

5.2 Introduction

An image can be thought of as a function from a compact domain (the "image plane") to the positive reals (the "intensity" range). Changes in the imaging conditions, such as in viewpoint and illumination, cause changes in both the domain and range of such a function. Also, a change of view of a Lambertian scene in ambient light can be modeled, away from occlusions, by a diffeomorphic deformation of the image domain [185]. Changes of illumination on a static scene can be modeled as structured changes in intensity (for instance, described by a low-dimensional linear variety, known as an "illumination cone"). Unfortunately, changes in the domain and range of the image play overlapping roles; one can always explain classes of images of the same scene or

"object" with changes in its domain (intensity values), modulo contrast functions [4], or by deformations of the image domain, as in a "deformable template" [71] (transitive actions of infinite-dimensional groups of diffeomorphisms). Therefore, inferring domain deformations and changes in intensity of a sequence of images obtained with different viewpoints and/or illumination is an ill-posed problem, and suitable regularizers have to be imposed to arrive at a meaningful model.

A common regularizer for changes in intensity is obtained by assuming that such changes cause the images to move on or close to a low-dimensional linear variety. The most common approach leads to principal components analysis (PCA), which has been used extensively in modeling and recognition of scenes when there are no changes of viewpoint [178]. Changes of viewpoint at a finite number of landmarks can also be modeled in a similar fashion, through procrustes analysis, which can be implemented by using singular value decompositions in a manner similar to PCA [63]. Combined changes in intensity and shape can be modeled in a conditionally linear fashion by assuming that intensity in a normalized frame is linear and that normalization is achieved by procrustes analysis, leading to so-called "active appearance models" [49]. These models have proven effective in modeling classes of objects, such as faces, with modest changes of appearance and shape when free of occlusions. Learning the principal components of shape and appearance from a collection of images of an object provides a powerful prior model that can be used to detect a new instance or to recognize the target class of an object.

The problem becomes significantly more complex in the presence of occlusions. In this case, domain deformations are not only not diffeomorphic, but they are not even regular functions since occlusions cause portions of the scene to disappear and other portions to appear. Jackson *et al.* [89] addressed the problem of modeling changes in motion and appearance of occlusion layers using the variational framework of deformotion introduced by Yezzi and Soatto [197], exploiting generic regularizers.

Using those methods in this chapter, a learning-based regularization approach is introduced that extends the active appearance model to scenes with occlusions, all within a principled variational framework. Since the shape, motion, and appearance of each layer are modeled, one can also fill in missing portions of layers (that are not visible in all images) thereby realizing a multi-view version of "image inpainting" [16]. For the case of just one layer, this model simplifies to standard active appearance models but is represented in a continuous domain rather than at a finite number of landmarks.

The shape of a layer that represents an object in the scene is transformed by a diffeomorphism to model small-scale changes in the object and a finite-dimensional group that describes coarse global motion. The intensity function associated with a layer is computed by minimizing a Mumford-Shah type energy that allows for occlusions and "in-paints" based on other images where there is no occlusion; if no information is available, then the solution of Laplace's equation is used. This chapter reduces the two most complex pieces, the intensity function for a layer and the diffeomorphism representing variations in shape, to a smaller, more reasonable space by using principal components analysis.

The model proposed describes changes in motion, appearance, depth ordering, and shape of a number of depth layers. In addition, a number of bases have to be learned for the appearance space, motion space, and shape space. This model is powerful, but the notation tends to become cumbersome when all factors are taken into account, and the computational cost of inference can be significant. For this reason, attention is restricted to the important case of two layers (foreground and background).

5.3 Layered Deformotion with a Joint Prior

A solution using layered deformotion would be difficult to get in scenes where the image statistics of each layer are similar. By fixing each layer Ω^k to an "average

shape," it becomes possible to examine the radiances ρ^k and the diffeomorphisms w^k and build a prior on them.

Having some prior models of the non-rigid registrations w_n^l and the smooth function ρ^l would greatly reduce the complexity of the above problem and allow for more accurate modeling. Using training data for specific objects results in faster and more accurate segmentations and registrations. The simplest reduction in computational complexity would be to represent the radiance ρ^l for layer l as

$$\rho^l(x) = \bar{\rho}^l(x) + \sum_{m=1}^{M} \alpha_m^l \rho_m^l(x), \forall x \in \Omega^l. \tag{34}$$

Principal components analysis will be used to get a good model of the radiance, so $\bar{\rho}^l(x)$ is the mean of that training data. The M principal components that represent the training data for the smooth function in layer l are written as $\rho_m^l(x)$.

The same procedure for learning the radiances will be performed to learn the deformations of an object. The deformations that map layer l into an image n decompose into a mean vector plus a weighted combination of principal components:

$$w_n^l(x) = \bar{w}^l(x) + \sum_{m=1}^{M} s_m^l w_m^l(x), \forall x \in \Omega^0 \tag{35}$$

.

A database of images of an object of interest is selected and the "layered deformotion" algorithm is run on them to "learn" the radiances and to "learn" the diffeomorphisms that object has in the database. The space of radiances and diffeomorphisms is reduced by using principal components analysis. Modeling the appearance and shape of any new object that belongs to a modeled class becomes much more accurate and computationally efficient.

5.4 Derivation of PCA flow with One Constant

Building $w(x)$ and $\rho^1(x)$ jointly out of principal components yields

$$[\rho^1(x), w(x)] = \left[\rho^1(x), x + u(x), y + v(x)\right] \tag{36}$$

and

$$[\rho^1(x), w(x)]=\left[\bar{\rho}^1(x)+\sum_{i=1}^{N} c_i\rho_i^1(x), x+\bar{u}(x)+\sum_{i=1}^{N} c_i u_i(x), y+\bar{v}(x)+\sum_{i=1}^{N} c_i v_i(x)\right] \quad (37)$$

where a bar over a variable denotes the mean. Recall the energy

$$\begin{aligned} E &= \int_{\hat{\Omega}^1} \left(I(\hat{x}) - \rho^1(x)\right)^2 - \left(I(\hat{x}) - \rho^0(\hat{x})\right)^2 d\hat{x} \\ &+ \beta \sum_{k=0}^{1} \int_{\Omega^k} \left\langle \nabla\rho^k, \nabla\rho^k \right\rangle dx \\ &+ \gamma \int_{\Omega^1} \langle \nabla u(x), \nabla u(x)\rangle + \langle \nabla v(x), \nabla v(x)\rangle \, dx \end{aligned} \quad (38)$$

and write it with principal components:

$$\begin{aligned} E &= \int_{\hat{\Omega}^1} (I(\hat{x}) - \rho^1(x,y))^2 - (I(\hat{x}) - \rho^0(\hat{x}))^2 d\hat{x} \\ &+ \beta \int_{\Omega^1} \left\langle \nabla\rho^1, \nabla \left(\bar{\rho}^1(x) + \sum_{i=1}^{N} c_i\rho_i^1(x)\right) \right\rangle dx \\ &+ \gamma \int_{\Omega^1} \left\langle \nabla u(x), \nabla \left(\bar{u}(x) + \sum_{i=1}^{N} c_i u_i(x)\right) \right\rangle dx \\ &+ \gamma \int_{\Omega^1} \left\langle \nabla v(x), \nabla \left(\bar{v}(x) + \sum_{i=1}^{N} c_i v_i(x)\right) \right\rangle dx \end{aligned} \quad (39)$$

where

$$\rho^1(x,y) = \rho^1(g^{-1}(\hat{x}) - \bar{u}(x) - \sum_{i=1}^{N} c_i u_i(x), g^{-1}(\hat{y}) - \bar{v}(x) - \sum_{i=1}^{N} c_i v_i(x)).$$

Differentiating yields:

$$\begin{aligned}
\frac{\partial E}{\partial c_j} &= \int_C \left\langle RS([u_j(x), v_j(x)]^T), J(g'w'T) \right\rangle f(x)\ ds \\
&+ 2\int_{\Omega^1} |g'||w'|(I(g(w(x))) - \rho^1(x)) \left\langle \nabla\rho^1(x), w_j(x) \right\rangle dx \\
&- 2\int_{\Omega^1} |g'||w'|(I(g(w(x))) - \rho^1(x))\rho_j^1(x)\ dx \qquad (40) \\
&+ 2\gamma \int_{\Omega^1} \langle \nabla u_j(x), \nabla \bar{u}(x) \rangle\, dx + 2\gamma \sum_{i=1}^{N} c_i \int_{\Omega^1} \langle \nabla u_j(x), \nabla u_i(x) \rangle\, dx \\
&+ 2\gamma \int_{\Omega^1} \langle \nabla v_j(x), \nabla \bar{v}(x) \rangle\, dx + 2\gamma \sum_{i=1}^{N} c_i \int_{\Omega^1} \langle \nabla v_j(x), \nabla v_i(x) \rangle\, dx \\
&+ 2\beta \int_{\Omega^1} \left\langle \nabla\rho_j^1(x), \nabla\bar{\rho}^1(x) \right\rangle dx + 2\beta \sum_{i=1}^{N} c_i \int_{\Omega^1} \left\langle \nabla\rho_j^1(x), \nabla\rho_i^1(x) \right\rangle dx,
\end{aligned}$$

where

$$f(x) = (I(g(w(x))) - \rho^1(x))^2 - (I(g(w(x))) - \rho^0(g(w(x))))^2. \qquad (41)$$

5.5 Experiments

The training set of 300 cars from the dataset from Jones *et al.*, [103] consisted of segmentations like that of Figure 23. An average shape was derived from all 300 examples, and each segmentation was registered (with a rigid component and a non-rigid component) to the average shape. Then, the image data could be mapped onto the average shape. From this process, the training data for the radiances ρ^1 and the warps w^1 was obtained. An example of the modes for the appearance is in Figure 24. An example mode for the diffeomorphisms is in Figure 25. Recall that for the warps there is a rigid component that captures variability in shape as well. An example of a joint mode for the appearance and the diffeomorphism is in Figure 26.

Using ten principal components for the radiance and the warp jointly, the results in Figures 27 and 28 were obtained. Figures 27 and 28 show the initial placement of the contour along with the segmentation obtained by using PCA for the radiances and the warps. The given example is not in the training database, as can be seen in

Figure 23: *Training example.*

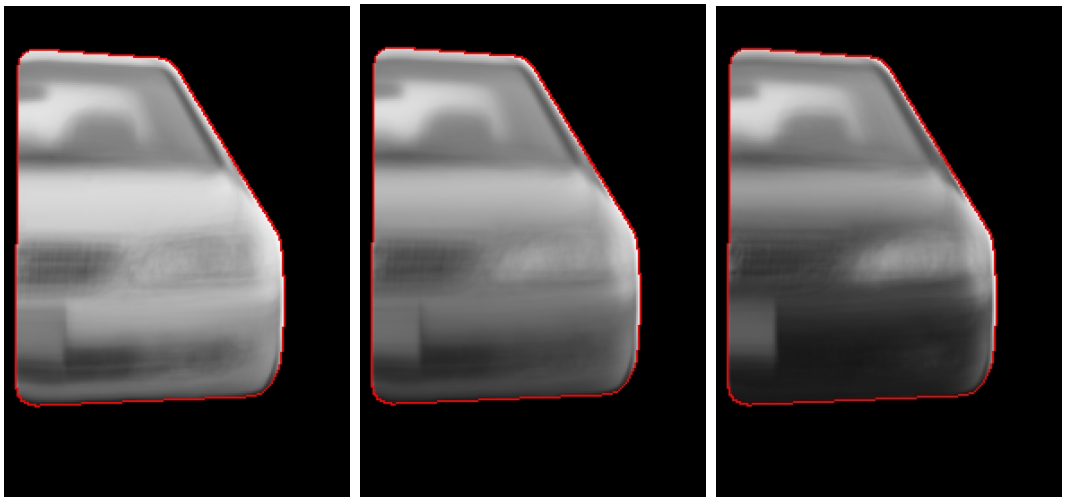

Figure 24: *Appearance modes: mean minus one standard deviation of the first mode; mean; mean plus one standard deviation of the first mode.*

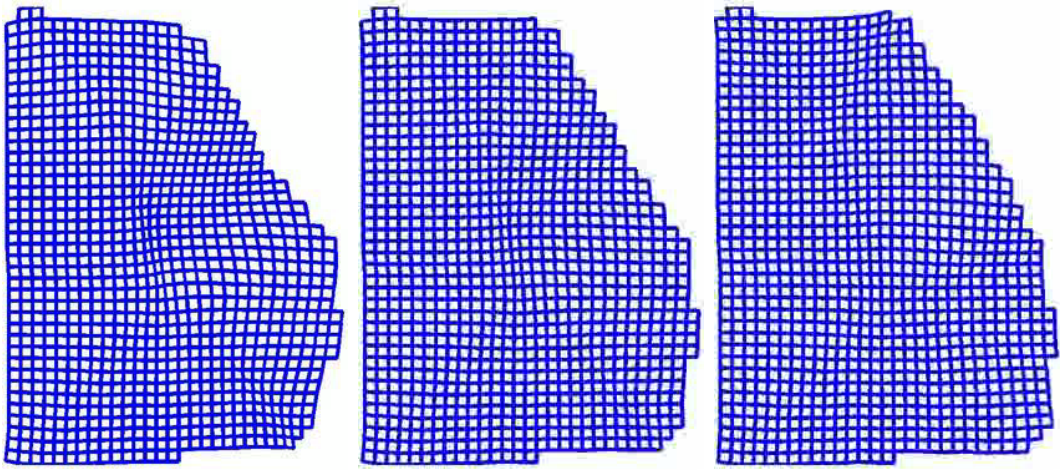

Figure 25: *Warp modes: varying the fourth mode.*

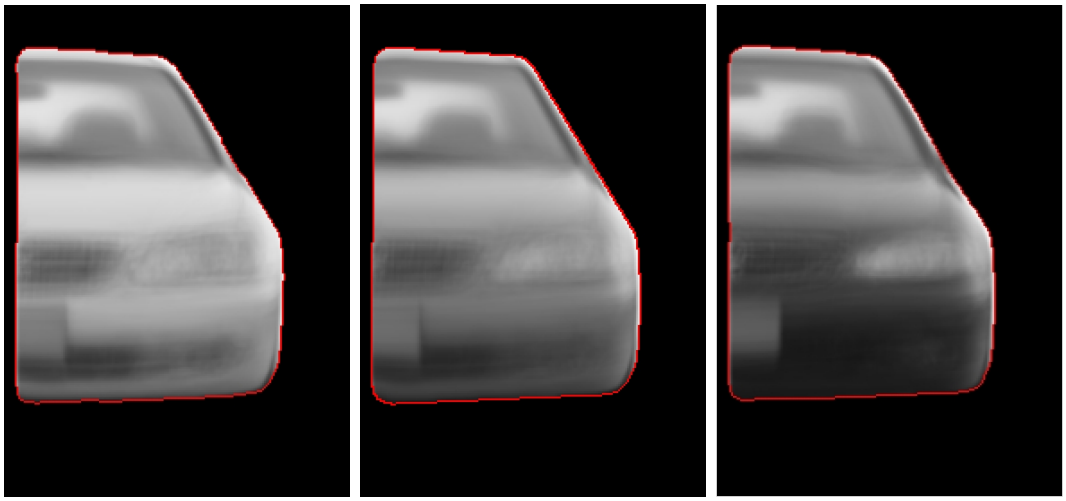

Figure 26: *Warp and appearance modes: mean minus one standard deviation of the first mode; mean; mean plus one standard deviation of the first mode.*

Figure 27: *Initialization; using the joint shape and appearance prior; PCA reconstruction; segmentation only a shape prior.*

Figure 28: *Initialization; using the joint shape and appearance prior; PCA reconstruction; segmentation using only a shape prior.*

the reconstruction of the radiance. The last pictures shown in both Figures 27 and 28 use only a prior on shape.

The benefits of training using shape and radiance to get a good segmentation are shown next. Results of segmenting the foreground object using a prior on both shape and radiance are compared to only using a prior on shape. In this example, the foreground object is a heart. Figure 29 is an example of the original image and the segmentation used for training. Sixteen datasets were used for training. The results in Figure 30 were obtained using ten principal components for the joint prior on the radiance and the warp. The first image in Figure 30 shows the initial placement of the curve, the second image shows the segmentation attained only using a shape prior,

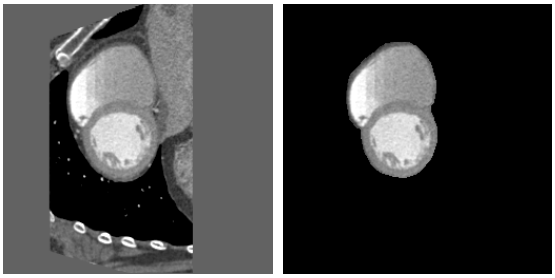

Figure 29: *Heart training set example: original image; hand segmentation.*

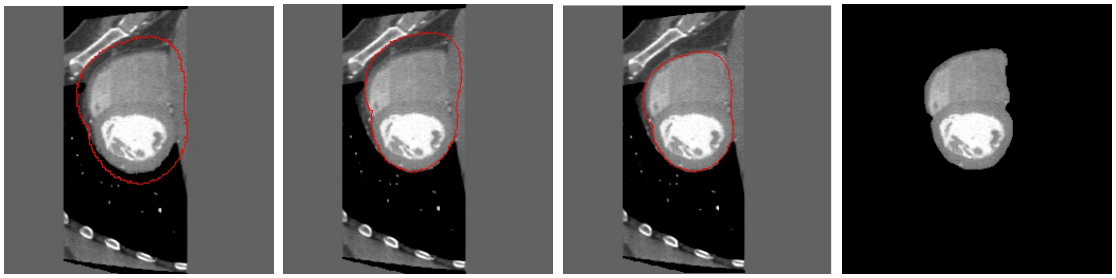

Figure 30: *Segmentation of a new heart image: initialization; segmentation using a prior on shape; segmentation using a prior on shape and radiance; hand segmentation.*

the third image shows the superior segmentation attained using the joint prior on shape and radiance, and the fourth image shows the actual hand segmentation (our ground truth).

5.6 Conclusion

Incorporating a joint prior on the appearance and shape for each layer significantly improves the method of layered deformotion. By fixing the shape of each layer, it becomes possible to perform dimensionality reduction (via principal components analysis) on the most complex functions for which to solve—the radiance functions and the warps. More useful and accurate segmentations are obtained by using the joint prior.

CHAPTER VI

CONCLUSION

This dissertation presented a model for the estimation of shape (image segmentation), appearance, pose (image registration), and movement (tracking) in a scene. This model used a separate layer to represent each component of the image. Each layer had three components: a contour that bounds the layer region, a smooth function that represents the appearance, and a transformation that maps that layer into an image. This method abstracts many existing methods in computer vision, such as L1 and L2 optical flow, motion segmentation, and image inpainting.

A segmentation method was presented for objects with regions of faint data, in particular, dendrites and their spines with their necks and heads. It incorporated a topology preservation method to include the areas with little or no data and to make sure the segmentation represents a single connected object. It used multiscale diffusion to coarsely segment the object, and then refined the segmentation by gradually diffusing less and less.

Next, a novel method created for tracking objects in a scene incorporated the segmenting curve as the state along with the parameters of a rigid registration that mapped that curve into each of the images. Using an observer framework, the method successfully tracked people and cars under severe occulsions.

Finally, layered deformotion with radiance was presented. The results (specifically, the smooth image-approximating function and the non-rigid registration) were taken from a database of images and were used to better model new images of those classes. This prior on both shape and appearance used principal components analysis on the smooth image-approximating function and the non-rigid registration and put this

information into the layered deformotion framework for quicker and more robust modeling. Results were shown using non-trivial background and foreground objects that were not included in the training set.

APPENDIX A

TOPOLOGY PRESERVATION

To explain how to preserve topology, one needs some background in digital topology to know what constitutes changes in topology of a surface in $\mathbb{Z}^3$. Some definitions from the paper by Bertrand [17] will be used. The surface will be a binary object with the pixels inside the surface being the foreground X and the pixels outside the surface being the background $\bar{X}$. The two measures of distance between two points $x = (x_1, x_2, x_3), y = (y_1, y_2, y_3)$ where $x, y \in \mathbb{Z}^3$ will be D_1 for the L_1 distance:

$$D_1(x, y) = \sum_{i=1}^{3} |y_i - x_i|, \tag{42}$$

and D_∞ for the L_∞ distance:

$$D_\infty(x, y) = \max_{i=1,2,3} |y_i - x_i|. \tag{43}$$

One also needs to define and build different types of the neighborhoods of a point x:

$$V_\infty^i(x) = \{y | D_\infty(x, y) \leq i\} \tag{44}$$

$$V_1^i(x) = \{y | D_1(x, y) \leq i\} \tag{45}$$

$$N_6^*(x) = V_1^1(x) \backslash \{x\} \tag{46}$$

$$N_{26}^*(x) = V_\infty^1(x) \backslash \{x\} \tag{47}$$

$$N_{18}^*(x) = \{V_1^2(x) \cup V_1^1(x)\} \backslash \{x\} \tag{48}$$

Within these neighborhoods of a point x, it is useful to know that the n-adjacency of the two points x, y is true when $y \in N_n^*(x)$. An n-path can then be formed from some point x_0 to x_k if for every x_i in that path (excluding x_0), x_i is n-adjacent to x_{i-1}. The existence of this n-path between any two points x and y means that they

are n-connected. To preserve the topology of a binary surface in $\mathbb{Z}^3$, the change of a point $x \in \mathbb{Z}^3$ from the foreground X to the background $\bar{X}$ (or vice versa) must not change the topology in a N^{26} neighborhood of itself. To check this, one needs to know if the point is a *simple point*. If it is, then one can change it from foreground to background or background to foreground.

A point x is a *simple point* if and only if there exists only two connected components in $N^*_{26}(x)$, one in the foreground and one in the background. To find if a point is simple, compute the number of connected components in $X \cap N^*_{26}(x)$ and the number of connected components in $\bar{X} \cap N^*_{18}(x)$. If they are both equal to one, x is a simple point. The assumption is that the point x is a point in the foreground. If x is a point in the background, the background is treated as X and the foreground is $\bar{X}$. For now, the case where x is a point in the foreground X, and the background is denoted by $\bar{X}$, the number of connected components in X and $\bar{X}$ (the topological numbers for a simple point) are denoted by

$$T_n(x, X) = 1 = T_{\tilde{n}}(x, \bar{X}), \tag{49}$$

where n and $\tilde{n}$ denote the connectivity used. In this case, $n = 26$ and $\tilde{n} = 6$ are used, so for the number of connected components in the foreground,

$$T_{26}(x, X) = \mathscr{C}^a_{26}[x, X \cap N^*_{26}(x)], \tag{50}$$

where $\mathscr{C}^a_{26}[x, X \cap N^*_{26}(x)]$ is the number of connected components in $X \cap N^*_{26}(x)$ using a 26-connectivity. For one connected component to exist in the foreground X, for every point $y \in X \cap N^*_{26}(x)$, y must be 26-connected to every other y, i.e. there must be a 26-path that connects all $y \in X \cap N^*_{26}(x)$. This computation of the number connected components is similar for the background, where the number of connected components is denoted as

$$T_6(x, X) = \mathscr{C}^a_6[x, \bar{X} \cap N^*_{18}(x)], \tag{51}$$

where $\mathscr{C}_6^a[x, \bar{X} \cap N_{18}^*(x)]$ is the number of connected components in $\bar{X} \cap N_{18}^*(x)$ using a 6-connectivity.

REFERENCES

[1] *2000 Conference on Computer Vision and Pattern Recognition (CVPR 2000), 13-15 June 2000, Hilton Head, SC, USA*, IEEE Computer Society, 2000.

[2] Allassonnière, S., A. Trouvé, and L. Younes, "Geodesic shooting and diffeomorphic matching via textured meshes.," in *EMMCVPR*, pp. 365–381, 2005.

[3] Alspach, D. and H. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Trans. on Automatic Control*, no. 17, pp. 438–448, 1972.

[4] Alvarez, L., F. Guichard, P. L. Lions, and J. M. Morel, "Axioms and fundamental equations of image processing," *Arch. Rational Mechanics*, vol. 123, 1993.

[5] Alvarez, L., J. Weickert, and J. Sanchez, "A scale-space approach to nonlocal optical flow calculations," in *In ScaleSpace '99*, pp. 235–246, 1999.

[6] Alvarez, L., J. Weickert, and J. Sanchez., "Reliable estimation of dense optical flow fields with large displacements," *IJCV*, vol. 39, pp. 41–56, Aug 2000.

[7] Alvino, C. V., *Multiscale Contour Methods in Computer Vision with Applications in Tomography*. PhD thesis, Georgia Institute of Technology, April 2005.

[8] Alvino, C. V. and A. J. Yezzi, "A scale space for contour registration using minimal surfaces.," in *Scale Space Methods in Computer Vision, 4th International Conference, Scale-Space 2003, Isle of Skye, UK, June 10-12, 2003, Proceedings* (Griffin, L. D. and M. Lillholm, eds.), vol. 2695 of *Lecture Notes in Computer Science*, pp. 164–179, Springer, 2003.

[9] Ayer, S. and H. S. Sawhney, "Layered representation of motion video using robust maximum-likelihood estimation of mixture models and mdl encoding," in *ICCV*, pp. 777–784, 1995.

[10] Azencott, R., F. Coldefy, and L. Younes, "A distance for elastic matching in object recognition," *Proc. 13th Intl. Conf. on Patt. Recog*, vol. 1, pp. 687–691, 1996.

[11] Baker, S., I. Matthews, and J. Schneider, "Image coding with active appearance models," tech. rep., Carnegie Mellon University, The Robotics Institute, 2003.

[12] Bar-Shalom, Y. and T. E. Fortmann, *Tracking and Data Association*. Academic Press Professional, Inc., 1987.

[13] Barron, J. L., D. J. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[14] Belhumeur, P. N., J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[15] Bergtholdt, M., D. Cremers, and C. Schnörr, "Variational segmentation with shape priors," in *Handbook of Mathematical Models in Computer Vision* (N. Paragios, Y. Chen, O. F., ed.), Springer, 2005.

[16] Bertalmio, M., G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of SIGGRAPH 2000, New Orleans, USA, July 2000*.

[17] Bertrand, G., "Simple points, toplogical numbers and geodesic neighborhoods in cubic grids," *Pattern Recognition Letters*, vol. 15, pp. 1003–1011, 1994.

[18] Black, M., "Recursive non-linear estimation of discontinuous flow fields," in *Proc. Third European Conf. on Computer Vision* (Eklundh, J., ed.), pp. 138–145, 1994.

[19] Black, M. and P. Anandan, "Robust dynamic motion estimation over time," in *Proc. Computer Vision and Pattern Recognition*, pp. 296–302, 1991.

[20] Black, M. J. and A. D. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation.," in Buxton and Cipolla [34], pp. 329–342.

[21] Blake, A. and R. Brockett., "On snakes and estimation theory," in *Submitted to the invited session on "Dynamic Vision" at the 33 Conf. on Decision and Control*, 1994.

[22] Blake, A. and M. Isard, *Active Contours*. Springer Verlag, 1998.

[23] Bookstein, L., "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 567–585, 1989.

[24] Boykov, Y. and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[25] Boykov, Y. and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images.," in *ICCV*, pp. 105–112, 2001.

[26] Boykov, Y. and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision.," in *EMMCVPR*, pp. 359–374, 2001.

[27] Boykov, Y. and V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts.," in *ICCV*, pp. 26–33, 2003.

[28] Boykov, Y. and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, 2004.

[29] Bregler, C., "Learning and recognizing human dynamics in video sequences.," in *Proc. of the Conference on Computer Vision and Pattern Recognition*, pp. 568–574, 1997.

[30] Breuß, M., T. Brox, A. Bürgel, T. Sonar, and J. Weickert, "Numerical aspects of TV flow," *Numerical Algorithms*, vol. 41, pp. 79–101, January 2006.

[31] Briggs, W. L., V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial, Second Edition*. SIAM, 2000.

[32] Brox, T., A. Bruhn, and J. Weickert, "Variational motion segmentation with level sets," in *ECCV*, 2006.

[33] Brox, T., A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *ECCV*, vol. 4, pp. 25–36, 2004.

[34] Buxton, B. F. and R. Cipolla, eds., *Computer Vision - ECCV'96, 4th European Conference on Computer Vision, Cambridge, UK, April 15-18, 1996, Proceedings, Volume I*, vol. 1064 of *Lecture Notes in Computer Science*, Springer, 1996.

[35] Cachier, P., E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, "Iconic feature based nonrigid registration: the pasha algorithm," *CVIU*, vol. 89, pp. 272–298, Feb-Mar 2003.

[36] Cevher, V. and J. H. McClellan, "General direction-of-arrival tracking using acoustic nodes," *Accepted for publication in IEEE Transactions on Signal Processing*, 2004.

[37] Chan, T. F. and L. A. Vese, "A level set algorithm for minimizing the mumford-shah functional in image processing," in *IEEE Proc. on Variational and Level Set Meth. in Comp. Vision*, pp. 161–168, 2001.

[38] Chan, T. and L. Vese, "Active contour without edges," *IEEE Transactions on Image Processing*, vol. 10, pp. 266–277, 2001.

[39] Chen, Y., H. D. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. A. Geiser, "Using prior shapes in geometric active contours in a variational framework," *IJCV*, vol. 50, no. 3, pp. 315–328, 2002.

[40] Cheng, Y., "Mean shift, mode seeking, and clustering.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, 1995.

[41] Cheung, V., B. J. Frey, and N. Jojic, "Video epitomes," in *CVPR*, vol. 1, pp. 42–49, 2005.

[42] Christensen, G. E., S. C. Joshi, and M. I. Miller, "Volumetric transformation of brain anatomy.," *IEEE Trans. Med. Imaging*, vol. 16, no. 6, pp. 864–877, 1997.

[43] Clarenz, U., M. Droske, S. Henn, M. Rumpf, and K. Witsch, "Computational methods for nonlinear image registration," 2005. to appear.

[44] Clarenz, U., M. Droske, and M. Rumpf, "Towards fast non–rigid registration," in *Inverse Problems, Image Analysis and Medical Imaging, AMS Special Session Interaction of Inverse Problems and Image Analysis*, vol. 313, pp. 67–84, AMS, 2002.

[45] Cohen, I., "Nonlinear variational method for optical flow computation," in *Proc. of the 8th Scandinavian Conference on Image Analysis*, May 1993.

[46] Cohen, L. D., "On active contour models and balloons," *CVGIP: Image Underst.*, vol. 53, no. 2, pp. 211–218, 1991.

[47] Cohen, L. D. and I. Cohen, "Finite-element methods for active contour models and balloons for 2-d and 3-d images.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1131–1147, 1993.

[48] Comaniciu, D. and P. Meer, "Mean shift: A robust approach toward feature space analysis.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[49] Cootes, T. F., G. J. Edwards, and C. J. Taylor), "Active appearance models," in *5th European Conference on Computer Vision (ECCV)* (Burkhardt, H. and B. Neumann, eds.), LNCS, pp. 484–498, 1998.

[50] Cox, I. J., "A review of statistical data association techniques for motion correspondence," *IJCV*, vol. 10, pp. 53–66, February 1993.

[51] Cremers, D., "Multiphase levelset framework for variational motion segmentation," in *Intl. Conf. on Scale-Space Theories in Computer Vision*, pp. 599–614, June 2003.

[52] Cremers, D., S. J. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: teaching level set to walk," in *Pattern Recognition, Proc. of DAGM, Springer LNCS Vol. 3175*, pp. 36–44, Tuebingen 2004.

[53] Cremers, D., S. J. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation," *International Journal of Computer Vision*, vol. 69, no. 3, pp. 335–351, 2006.

[54] Cremers, D. and S. Soatto, "Motion competition: A variational framework for piecewise parametric motion segmentation," *International Journal of Computer Vision*, vol. 62, pp. 249–265, May 2005.

[55] Cremers, D., F. Tischhäuser, J. Weickert, and C. Schnörr, "Diffusion snakes: Introducing statistical shape knowledge into the Mumford–Shah functional," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 295–313, 2002.

[56] Davies, R., C. Twining, T. Cootes, J. Waterton, and C. Taylor, "3d statistical shape models using direct optimisation of description length," in *7th European Conference on Computer Vision (ECCV)*, vol. 3, pp. 3–21, 2002.

[57] Deriche, R., P. Kornprobst, and G. Aubert, "Optical flow estimation while preserving its discontinuities: A variational approach," vol. 2, pp. 71–80, Dec. 1995.

[58] Doretto, G., D. Cremers, P. Favaro, and S. Soatto, "Dynamic texture segmentation," in *Intl. Conf. on Comp. Vision*, pp. (2) 1236–1242, October 2003.

[59] Doretto, G., E. Jones, and S. Soatto, "Spatially homogeneous dynamic textures," in *Proc. of the Eur. Conf. on Comp. Vision*, pp. 591–602, May 2004.

[60] Doretto, G. and S. Soatto, "Towards plenoptic dynamic textures," in *Proc. of the Workshop Textures 2003*, pp. 25–30, October 2003.

[61] Doretto, G. and S. Soatto, "Modeling dynamic scenes with active appearance," in *Technical Report UCLA-CSD 04-0053*, December 2004.

[62] Doretto, G. and S. Soatto, *O. Faugeras and N. Paragios (Ed.)*, ch. Dynamic Textures. Springer Verlag, 2005.

[63] Dryden, I. L. and K. V. Mardia, *Statistical Shape Analysis*. Chichester: Wiley, 1998.

[64] Duci, A., A. Yezzi, S. Mitter, and S. Soatto, "Region matching with missing parts," in *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, vol. 3, pp. 28–64, 2002.

[65] Duci, A., A. J. Yezzi, S. K. Mitter, and S. Soatto, "Shape representation via harmonic embedding," in *Intl. Conf. on Comp. Vision*, pp. 656–662, October 2003.

[66] Duci, A., A. J. Yezzi, S. K. Mitter, and S. Soatto, "Region matching with missing parts," *Image and Vision Computing*, vol. 24, pp. 271–277, March 2006.

[67] Elgammal, A. M., R. Duraiswami, and L. S. Davis, "Efficient kernel density estimation using the fast gauss transform with applications to color modeling and tracking.," *PAMI*, vol. 25, no. 11, pp. 1499–1504, 2003.

[68] Frey, B. J., N. Jojic, and A. Kannan, "Learning appearance and transparency manifolds of occluded objects in layers," in *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR '03)*, pp. 45–52, 2003.

[69] Goldenberg, R., R. Kimmel, E. Rivlin, and M. Rudzsky, "Fast geodesic active contours," *IEEE Trans. on Image Processing*, vol. 10(10), pp. 1467–75, October 2001.

[70] Goshtasby, A., L. Staib, C. Studholme, and D. Terzopoulos, "Nonrigid image registration: guest editors' introduction," *Computer Vision and Image Understanding*, vol. 89, pp. 109–113, Feb-Mar 2003.

[71] Grenander, U., *Lectures in Pattern Theory*. Berlin: Springer, 1976.

[72] Grenander, U., *General Pattern Theory*. Oxford University Press, 1993.

[73] Grenander, U. and M. Miller, "Representations of knowledge in complex systems," *Journal of the Royal Statistical Society B*, vol. 56, pp. 249–603, 1993.

[74] Griffin, L. D. and M. Lillholm, eds., *Scale Space Methods in Computer Vision, 4th International Conference, Scale-Space 2003, Isle of Skye, UK, June 10-12, 2003, Proceedings*, vol. 2695 of *Lecture Notes in Computer Science*, Springer, 2003.

[75] Haker, S., L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *IJCV*, vol. 60, no. 3, pp. 225–240, 2004.

[76] Han, X., C. Xu, D. Tosun, and J. L. Prince, "Corical surface reconstruction using a topology preserving geometric model.," *IEEE Trans. on Medical Imaging*, pp. 109–121, 2002.

[77] Haussecker, H. W. and D. J. Fleet, "Computing optical flow with physical models of brightness variation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23(6), pp. 661–673, 2001.

[78] Horn, B. K. P. and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[79] Hsu, S., P. Anandan, and S. Peleg, "Accurate computation of optical flow by using layered motion representations," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, pp. 743–746, Oct 1994.

[80] Huang, X., Z. Li, and D. Metaxas, "Learning coupled prior shape and appearance models for segmentation," in *MICCAI*, no. 3216 in LNCS, pp. 60–69, 2004.

[81] Huang, X., D. Metaxas, and T. Chen, "Metamorphs: Deformable shape and texture models," in *CVPR*, vol. I, pp. 496–503, 2004.

[82] Huang, X., N. Paragios, and D. Metaxas, "Registration of structures in arbitrary dimension: Implicit representations, mutual information and free form deformations," Tech. Rep. DCS-TR-520, Division of Computer and Information Sciences, Rutgers University, New Brunswick, April 2003.

[83] Huang, X., N. Paragios, and D. Metaxas, "Shape registration in implicit spaces using information theory and free form deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2006.

[84] Huang, X., Z. Qian, R. Huang, and D. Metaxas, "Deformable-model based textured object segmentation," in *EMMCVPR*, no. 3757 in LNCS, pp. 119–135, Springer, 2005.

[85] Isard, M. and A. Blake, "Condensation – conditional density propagation for visual tracking," 1998.

[86] Isard, M., *Visual Motion Analysis by Probabilistic Propagation of Conditional Density*. PhD thesis, Oxford University, 1998.

[87] Isard, M. and A. Blake, "Contour tracking by stochastic propagation of conditional density.," in *ECCV (1)*, pp. 343–356, 1996.

[88] .Jackson, J. D., A. Yezzi, and S. Soatto, "Tracking deformable objects under severe occlusions," in *43rd IEEE Conference on Decision and Control*, 2004.

[89] Jackson, J. D., A. Yezzi, and S. Soatto, "Dynamic shape and appearance modeling via moving and deforming layers," in *EMMCVPR 2005*, 2005.

[90] Jackson, J. D., A. J. Yezzi, W. Wallace, and M. F. Bear, "Segmentation of coarse and fine scale features using multi-scale diffusion and mumford-shah," in *Scale Space Methods in Computer Vision, 4th International Conference, Scale-Space 2003, Isle of Skye, UK, June 10-12, 2003, Proceedings* (Griffin, L. D. and M. Lillholm, eds.), vol. 2695 of *Lecture Notes in Computer Science*, pp. 615–624, Springer, 2003.

[91] Jazwinski, A., *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[92] Jin, H., D. Cremers, A. Yezzi, and S. Soatto, "Shedding light in stereoscopic segmentation," in *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, July 2004.

[93] Jin, H., S. Soatto, and A. Yezzi, "Multi-view stereo beyond lambert," in *Conf. on Computer Vision and Pattern Recognition*, June 2003.

[94] Jin, H., S. Soatto, and A. Yezzi, "Multi-view stereo reconstruction of dense shape and complex appearance," Tech. Rep. 04-53, UCLA, 2004.

[95] Jin, H., S. Soatto, and A. Yezzi, "Multi-view stereo reconstruction of dense shape and complex appearance," *Intl. J. Computer Vision*, vol. 63, pp. 175–189, July 2005.

[96] Jin, H., R. Tsai, L. Chen, A. Yezzi, and S. Soatto, "Estimation of 3d surface shape and smooth radiance from 2d images: A level set approach," *J. of Sci. Comp.*, vol. 19(1-3), pp. 267–292, 2003.

[97] Jin, H., A. Yezzi, and S. Soatto, "Variational multiframe stereo in the presence of specular reflections," in *Proc. of the Intl. Conf. on 3DPVT*, pp. 626–630, June 2002.

[98] Jin, H., A. Yezzi, and S. Soatto, "Stereoscopic shading: integrating shape cues in a variational framework," in *Intl. Conf. on Computer Vision and Pattern Recognition*, pp. 169–176, June 2000.

[99] Jin, H., A. J. Yezzi, and S. Soatto, "Region-based segmentation on evolving surfaces with application to 3d shape and radiance estimation," in *Proc. of the Eur. Conf. on Comp. Vision*, pp. 114–125, May 2004.

[100] Jin, H., *Variational Methods for Shape Reconstruction in Computer Vision*. PhD thesis, Washington University, August 2003.

[101] Jojic, N. and B. J. Frey, "Learning flexible sprites in video layers," in *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 199–206, 2001.

[102] Jojic, N., B. J. Frey, and A. Kannan, "Epitomic analysis of appearance and shape," in *ICCV*, vol. 1, pp. 34–43, 2003.

[103] Jones, E. and S. Soatto, "Layered active appearance models," in *ICCV*, vol. 2, pp. 1097–1102, Oct 2005.

[104] Kalman, R., "A new approach to linear filtering and prediction problems," *Transaction of the ASME–Journal of Basic Engineering*, pp. 35–45, Mar. 1960.

[105] Kass, M., A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proc. of IEEE Conference on Computer Vision*, (London, England), pp. 259–268, 8-11 1987.

[106] Kass, M., A. Witkin, and D. Terzopoulos, "Snakes: Active shape models," *International Journal of Computer Vision*, vol. 1, pp. 321–331, 1987.

[107] Kenigsberg, A., R. Kimmel, and I. Yavneh, "A multigrid approach for fast geodesic active contours," Tech. Rep. CIS-2004-06 - 2004, 2004.

[108] Kichenassamy, S., A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi, "Gradient flows and geometric active contour models.," in *ICCV*, pp. 810–815, 1995.

[109] Kimia, B., A. Tannenbaum, and S. Zucker, "Shapes, shocks, and deformations i: the components of two-dimensional shape and the reaction-diffusion space.," *International Journal of Computer Vision*, vol. 15, pp. 189–224, 1995.

[110] Kimmel, R. and A. Bruckstein, "Tracking level sets by level sets: a method for solving the shape from shading problem.," *Computer Vision, Graphics and Image Understanding*, vol. 62, no. 1, pp. 47–58, 1995.

[111] Kimmel, R., N. Kiryati, and A. M. Bruckstein, "Multivalued distance maps for motion planning on surfaces with moving obstacles," *IEEE Transactions on Robotics and Automation*, vol. 14, pp. 427–436, June 1998.

[112] Kolmogorov, V. and Y. Boykov, "What metrics can be approximated by geo-cuts, or global optimization of length/area and flux.," in *ICCV*, pp. 564–571, 2005.

[113] Kuhne, G., J. Weickert, M. Beier, and W. Effelsberg, "Fast implicit active contour models.," in *Pattern Recognition, 24th DAGM Symposium, Zurich,Switzerland, September 16-18, 2002, Proceedings*, vol. 2449 of *Lecture Notes in Computer Science*, pp. 133–140.

[114] Kumar, A., A. R. Tannenbaum, and G. J. Balas, "Optical flow: a curve evolution approach.," *IEEE Transactions on Image Processing*, vol. 5, no. 4, pp. 598–610, 1996.

[115] Leroy, B., I. Herlin, and L. D. Cohen, "Multi-resolution algorithms for active contour models," in *Proc. 12th International Conference on Analysis and Optimization of Systems: Images, Wavelets and PDE's (ICAOS'96)*, 1996.

[116] Leventon, M. E., W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," *CVPR*, vol. 01, pp. 316–323, 2000.

[117] Lucas, B. and T. Kanade, "An iterative image registration technique with application to stereo vision," in *Proc. DARPA IU Workshop*, pp. 121–130, 1981.

[118] Ma, Y., S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3-D Vision*, ch. 2. Springer, 2003.

[119] Maintz, J. and M. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[120] Malandain, G. and G. Bertrand, "Fast characterization of 3d simple points," *IEEE Pattern Recognition*, pp. 232–235, 1992.

[121] Malandain, G. and G. Bertrand, "A new characterization of three-dimensional simple points," *Pattern Recognition Letters*, vol. 15, pp. 169–175, 1994.

[122] Manay, S., A. J. Yezzi, B. W. Hong, and S. Soatto, "Integral invariant signatures," in *Proc. of the Eur. Conf. on Comp. Vision*, pp. 87–94, May 2004.

[123] Manay, S., *Applications of Anti-Geometric Diffusion to Computer Vision:Thresholding, Segmentation, and Distance Functions*. PhD thesis, Georgia Institute of Technology, April 2003.

[124] Matthews, I. and S. Baker, "Active appearance models revisted," *IJCV*, vol. 60, no. 2, pp. 135–164, 2004.

[125] Meltzer, J., M. Yang, R. Gupta, and S. Soatto, "Multiple view feature descriptors from image sequences via kernel principal component analysis," in *Proc. of the Eur. Conf. on Comp. Vision*, pp. 215–227, May 2004.

[126] Mennucci, A. C. G., A. Yezzi, and G. Sundaramoorthi, "Sobolev–type metrics in the space of curves," 2006.

[127] Metaxas, D., X. Huang, and T. Chen, *Mathematical Models in Computer Vision: The Handbook*, ch. Integrating Shape and Texture in Deformable Models: From Hybrid Methods to Metamorphs. Springer, 2005.

[128] Miller, M. I. and L. Younes, "Group action, diffeomorphism and matching: a general framework," in *Proc. of SCTV*, 1999.

[129] Moelich, M. and T. Chan., "Tracking objects with the chan-vese model," tech. rep., UCLA CAM Report, 2003.

[130] Mumford, D. and J. Shah, "Optimal approximations .," *J. of Comp. Physics*, vol. 79, pp. 12–49, 1988.

[131] Nagel, H. and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, 1986.

[132] Niethammer, M. and A. Tannenbaum, "Dynamic geodesic snakes for visual tracking," in *CVPR*, 2004.

[133] Nimchinsky, E. A., B. Sabatini, and K. Svoboda., "Structure and function of dendritic spines," *Annu. Rev. Physiol.*, pp. 313–353, 2002.

[134] Osher, S. and J. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi equations.," *J. of Comp. Physics*, vol. 79, pp. 12–49, 1988.

[135] Papenberg, N., A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping," *International Journal of Computer Vision*, vol. 67, pp. 141–158, April 2006.

[136] Papoulis, A. and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. Boston, MA: McGraw-Hill, fourth ed., 2002.

[137] Paragios, N. and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 266–280, Mar. 2000.

[138] Paragios, N. and R. Deriche, "Geodesic active regions and level set methods for motion estimation and tracking," *Computer Vision and Image Understanding*, vol. 97, no. 3, pp. 259–282, 2005.

[139] Paragios, N. and R. Deriche, "Geodesic active regions for motion estimation and tracking.," in *ICCV*, pp. 688–694, 1999.

[140] Paragios, N. and R. Deriche, "Geodesic active regions for supervised texture segmentation.," in *ICCV*, pp. 926–932, 1999.

[141] Paragios, N. and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *International Journal of Computer Vision*, vol. 46, pp. 223–247, February 2002.

[142] Paragios, N., O. Mellina-Gottardo, and V. Ramesh, "Gradient vector flow fast geodesic active contours.," in *ICCV*, pp. 67–75, 2001.

[143] Paragios, N., M. Rousson, and V. Ramesh, "Non-rigid registration using distance functions," *Comput. Vis. Image Underst.*, vol. 89, no. 2-3, pp. 142–165, 2003.

[144] Paragios, N., M. Taron, X. Huang, M. Rousson, and D. Metaxas, *Statistics and Analysis of Shapes*, ch. On the Representation of Shapes Using Implicit Functions. Springer, 2005.

[145] Perona, P. and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

[146] Peterfreund, N., "Robust tracking of position and velocity with kalman snakes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 6, pp. 564–569, 1999.

[147] Peterfreund, N., "The velocity snake: deformable contour for tracking in spatio-velocity space," *Comput. Vis. Image Underst.*, vol. 73, no. 3, pp. 346–356, 1999.

[148] Pluim, J., J. B. Maintz, and M. Viergever, "Mutual information based registration of medical images: a survey," *IEEE Trans. on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[149] Rasmussen, C. and G. Hager, "Probabilistic data association methods for tracking complex visual objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 560–576, 2001.

[150] Rathi, Y., N. Vaswani, A. Tannenbaum, and A. Yezzi, "Particle filtering for geometric active contours with application to tracking moving and deforming objects," in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005*, 2005.

[151] Ronfard, R., "Region-based strategies for active contour models.," *Int. J. Comput. Vis.*, vol. 13, pp. 229–251, 1994.

[152] Rueckert, D., A. Frangi, and J. Schnabel, "Automatic construction of 3-d statistical deformation models of the brain using non-rigid registration," *IEEE Trans. on Medical Imaging*, vol. 22, pp. 1014–1025, Aug. 2003.

[153] Rueckert, D., A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3d statistical deformation models of the brain using non-rigid registration.," *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 1014–1025, 2003.

[154] Rueckert, D., L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Non-rigid registration using free-form deformations: Application to breast mr images.," *IEEE Trans. Med. Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[155] Rumpf, M. and R. Strzodka, "Using graphics cards for quantized FEM computations," in *Proceedings of IASTED Visualization, Imaging and Image Processing Conference (VIIP'01)*, pp. 193–202, 2001.

[156] Samson, C., L. Blanc-Féraud, G. Aubert, and J. Zerubia, "A level set model for image classification.," *International Journal of Computer Vision*, vol. 40, no. 3, pp. 187–197, 2000.

[157] Schnabel, J. A., D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations.," in *MICCAI*, pp. 573–581, 2001.

[158] Schnörr, C., "Computation of discontinuous optical flow by domain decomposition and shape optimization," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 153–165, 1992.

[159] Sethian, J., *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Material Science.* Cambridge, U.K.: Cambridge University Press, 1999.

[160] Soatto, S., G. Doretto, and Y. Wu, "Dynamic textures," in *Proc. of the Intl. Conf. on Computer Vision*, pp. 439–446, 2001.

[161] Soatto, S., A. J. Yezzi, and A. Duci, *Geometric Level Set Methods in Imaging, Vision and Graphics, S. J. Osher and N. Paragios (Eds.)*, ch. Region matching and tracking under deformations and occlusions, pp. 319–340. Springer Verlag, 2003.

[162] Soatto, S., A. J. Yezzi, and H. Jin, "Tales of shape and radiance in multiview stereo," in *Intl. Conf. on Comp. Vision*, pp. 974–981, October 2003.

[163] Strzodka, R., M. Droske, and M. Rumpf, "Fast image registration in DX9 graphics hardware," *Journal of Medical Informatics and Technologies*, vol. 6, pp. 43–49, Nov 2003.

[164] Sundaramoorthi, G., J. D. Jackson, A. Yezzi, and A. Mennucci, "Tracking with sobolev active contours," in *CVPR*, 2006.

[165] Sundaramoorthi, G. and A. Yezzi, "Global regularizing flows with topology preservation for active contours and polygons," tech. rep., Georgia Institute of Technology, Aug 2005.

[166] Sundaramoorthi, G. and A. Yezzi, "More-than-topology preserving flows for active contours and polygons," in *ICCV*, Oct 2005.

[167] Sundaramoorthi, G. and A. Yezzi, "Global regularizing flows with topology preservation for active contours and polygons," *IEEE Transactions on Image Processing*, to appear.

[168] Sundaramoorthi, G., A. Yezzi, and A. Mennucci, "Sobolev active contours," in *VLSM (in conjunction with ICCV)*, 2005.

[169] Sundaramoorthi, G., A. Yezzi, and A. Mennucci, "Sobolev active contours," tech. rep., Georgia Institute of Technology, Dec 2005.

[170] Szeliski, R. and J. Coughlan, "Hierarchical spline-based image registration," in *CVPR*, pp. 194–201, June 1994.

[171] Szeliski, R. and J. Coughlan, "Spline-based image registration," Tech. Rep. 94/1, Digital Equipment Corporation, Cambridge Research Lab, April 1994.

[172] Szeliski, R. and J. Coughlan, "Spline-based image registration," *International Journal of Computer Vision*, vol. 22, pp. 199–218, March/April 1997.

[173] Terzopoulos, D. and R. Szeliski, *Active Vision*, ch. Tracking with Kalman Snakes, pp. 3–20. MIT Press, 1992.

[174] Trouvé, A. and L. Younes, "Metamorphoses through lie group action.," *Foundations of Computational Mathematics*, vol. 5, no. 2, pp. 173–198, 2005.

[175] Tsai, A., A. Y. Jr., W. W. III, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "Model-based curve evolution technique for image segmentation," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 463–468.

[176] Tsai, A., A. Yezzi, and A. Wilsky, "Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification.," *IEEE Trans. on Image Processing*, pp. 1169–1184, 2001.

[177] Turk, M., *Interactive Time Vision: Face Recognition as a Visual Behavior*. PhD thesis, MIT, September 1991.

[178] Turk, M. and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[179] Turk, M. and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1991.

[180] Tuzel, O. and P. Meer, "Mean shift clustering.".

[181] Unal, G., A. Yezzi, and H. Krim, "Active polygons for object tracking," in *First International Symposium 3D Data Processing, Visualization, Transmission*, June 2002. Padova, Italy.

[182] Unal, G., A. Yezzi, and H. Krim, "A vertex-based representation of objects in an image," in *Proc. of Int. Conf. Image Processing*, pp. 896–899, Sept. 2002.

[183] Unal, G., *Curve and Polygon Evolution Techniques for Image Processing*. PhD thesis, North Carolina State University, 2002.

[184] V. Caselles, R. Kimmel, G. S., "Geodesic active contours," in *Proceedings of the ICCV*, 1995.

[185] Vedaldi, A. and S. Soatto, "Features for recognition: viewpoint invariance for non-planar scenes," in *Technical Report UCLA-CSD 04-0049*, December 2004.

[186] Viola, P. A., "Alignment by maximization of mutual information," Tech. Rep. AITR-1548, 1995.

[187] Wang, J. Y. A. and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. on Image Processing*, vol. 3, pp. 625–638, Sep. 1994.

[188] Weickert, J. and G. Kuhne, *Geometric Level Set Methods in Imaging, Vision, and Graphics*, ch. Fast methods for implicit active contour models, pp. 43–58. New York: Springer, 2003.

[189] Weickert, J. and C. Schnörr, "Variational optic flow computation with a spatio-temporal smoothness constraint," *Journal of Mathematical Imaging and Vision*, vol. 14, pp. 245–255, May 2001.

[190] Weickert, J., B. M. ter Haar Romeny, and M. A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Trans. on Image Processing*, vol. 7, pp. 398–410, March 1998.

[191] Weinstock, R., *Calculus of Variations: With Applications to Physics and Engineering*. New York: Dover Pub. Inc., 1974.

[192] Welch, G. and G. Bishop, "An introduction to the kalman filter," *Technical Report, Department of Comp. Sc. and Engg., Univ. of North Carolina at Chapel Hill*, Mar. 2002.

[193] Xu, C. and J. L. Prince, "Gradient vector flow: A new external force for snakes," in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, p. 66, IEEE Computer Society, 1997.

[194] Xu, C. and J. L. Prince, "Snakes, shapes, and gradient vector flow.," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.

[195] Yezzi, A., "Modified curvature motion for image smoothing and enhancement," *IEEE Trans. Image Processing*, pp. 345–352, 1998.

[196] Yezzi, A. and S. Soatto, "Stereoscopic segmentation," in *Proc. of the Intl. Conf. on Computer Vision*, pp. 59–66, 2001.

[197] Yezzi, A. and S. Soatto, "Deformotion: deforming motion, shape average and the joint segmentation and registration of images.," *Intl. J. of Comp. Vis.*, vol. 53, no. 2, pp. 153–167, 2003.

[198] Yezzi, A., S. Soatto, A. Tsai, and A. Willsky, *Mathematics and Multimedia*, ch. Curve and Surface Evolution for Image Segmentation and Stereo Reconstruction Using the Mumford-Shah Functional. IMA, 2002.

[199] Yezzi, A., L. Zollei, and T. Kapur, "A variational framework for joint segmentation and registration," in *Mathematical Methods in Biomedical Image Analysis*, pp. 44–51, Dec 2001.

[200] Yezzi, A., L. Zollei, and T. Kapur, "A variational framework for integrating segmentation and registration through active contours," *Medical Image Analysis*, vol. 7, pp. 171–185, June 2003.

[201] Yezzi, A. J., S. Soatto, H. Jin, A. Tsai, and A. Willsky, *Geometric Level Set Methods in Imaging, Vision and Graphics, S. J. Osher and N. Paragios (Eds.)*, ch. Mumford-Shah from segmentation to stereo, pp. 207–228. Springer Verlag, 2003.

[202] Yezzi, A. J., S. Kichenassamy, A. Kumar, P. J. Olver, and A. Tannenbaum, "Active contours for visual tracking: a geometric gradient based approach," in *Proceedings of the 34th IEEE Conference on Decision and Control*, pp. 4041–4046, 1995.

[203] Yezzi, A. J., S. Kichenassamy, A. Kumar, P. J. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery.," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 199–209, 1997.

[204] Yezzi, A. J., A. Tsai, and A. S. Willsky, "Binary and ternary flows for image segmentation.," in *ICIP (2)*, pp. 1–5, 1999.

[205] Yezzi, A. J., A. Tsai, and A. S. Willsky, "A statistical approach to snakes for bimodal and trimodal imagery.," in *ICCV*, pp. 898–903, 1999.

[206] Younes, L., "Computable elastic distances between shapes," *SIAM J. of Appl. Math.*, 1998.

[207] Yuille, A. L., "Deformable template for face recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 59–70, 1991.

[208] Yuille, A. L., P. W. Hallman, and D. S. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, pp. 99–111, 1992.

[209] Zaritsky, R., N. Peterfreund, and N. Shimkin, "Velocity-guided tracking of deformable contours in three dimensional space," *Int. J. Comput. Vision*, vol. 51, no. 3, pp. 219–238, 2003.

[210] Zhang, R., N. Paragios, and D. Metaxas, "Implicit representations for recovery and reconstruction of motion layers," in *2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision, organized in Conjunction with the 9th IEEE International Conference in Computer Vision*, 2003.

[211] Zhu, S. C., T. S. Lee, and A. L. Yuille, "Region competition: Unifying snakes, region growing, energy/bayes/MDL for multi-band image segmentation," in *ICCV*, pp. 416–423, 1995.

[212] Zhu, S. C. and A. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, 1996.

VITA

Jeremy Jackson was born in Dallas, Texas in 1974. He attended high school at Germantown High School in Germantown, Tennessee. He studied trumpet performance under the direction of Mr. Thomas Booth at Southern Methodist University and received a Bachelors of Music in 1996. Going into engineering, he attended Tulane University where he graduated Summa cum Laude with a double degree in Electrical Engineering and Computer Science in 2001. While at Tulane, he was a teaching assistant for Digital Logic and Computer Organization and taught Sparc Assembly. During his senior year, he taught himself computer vision techniques and implemented a face tracking program for his senior project.

In Fall 2001, he joined the Ph.D. program at the Georgia Institute of Technology and was awarded a President's Fellowship. His Ph.D. research was guided by Professor Anthony J. Yezzi in the field of computer vision using partial differential equations. His research has dealt with image segmentation, registration, tracking, and appearance modeling. He is a member of IEEE and Eta Kappa Nu. He has worked with Tom Malzbender at Hewlett Packard doing work in the research of image relighting of scenes. He was awarded a Ph.D. in August 2007 from the School of Electrical and Computer Engineering at the Georgia Institute of Technology with a minor in Mathematics.